



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 117719

TO: Celine Qian
Location: rem/2a89/2c70
Art Unit: 1636
Thursday, April 01, 2004

Case Serial Number: 09/913878

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart



STIC SEARCH RESULT FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



BEST AVAILABLE COPY

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

ATTN: Ed Hart

Access DB# 117719

SEARCH REQUEST FORM

CRFE

Scientific and Technical Information Center

Requester's Full Name Debra Giam Examiner # 78710 Date 3/25/04
Alt Unit 1636 Phone Number 306-0183 Serial Number 07/13878
Mail Box and Bldg Room Location 2A89 Results Format Preferred (circle) PAPER ~~DISK~~ ~~E-MAIL~~

If more than one search is submitted, please prioritize searches in order of need. *mej*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention Isolation & characterization of an Orna silencing gene

Inventors (please provide full names): Biseppe, Hanne

Earliest Priority Filing Date 11/6/2000

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search ☒ SEQ ID NO:1 NA-8045

☒ 2) nucleic acid encoding seq ID NO:2 AA-1402

☒ 3) aa 710 - aa 1282 of seq ID NO:2

☒ 4) nucleic acid encoding 710-1282 of seq ID NO:2

mej

STAFF USE ONLY

Searcher _____

Searcher Phone # _____

Searcher Location _____

Date Searched 3/29/04

Date Completed _____

Searcher Prep & Review Time _____

Cost Prep Time _____

Cost Time _____

Type of Search

NA Sequence (#) 3

AA Sequence (#) 2+1

Structure (#) _____

Bibliographic _____

Citation _____

Full text _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Bridge _____

Quickie _____

Online _____

Lexis Nexis _____

Sequence Systems _____

RAW Internet _____

Other _____

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 02:06:52 ; Search time 3597.57 Seconds
(without alignments)
6903.427 Million cell updates/sec

Title: US-09-913-8/8a-2_COPY_710_1282
Perfect score: 3010
Sequence: 1 GVGRMSRSVAKRIEDVLGLG.....EKWCAITPEAMDKSGANYDS 573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODBL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTC_spool/US09913878/runat_29032004_082008_8366/app_query_fasta_1.2254
-DB=GenEmbl -QMT=frascap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORES=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878@cgn.1.1.7684@runat_29032004_082008_8366 -NCPU=6 -ICPU=3
-NO_MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6
-PGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_cm.*
21: em_ov.*
22: em_gv.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|------------------------------|
| 1 | 301.0 | 100.0 | 4206 | 8 | NCR133528 Neurospor |
| 2 | 301.0 | 100.0 | 8045 | 6 | AX034425 Sequence |
| 3 | 275.5 | 9.2 | 3505 | 8 | NTA011576 Nicotiana |
| 4 | 268.5 | 8.9 | 3807 | 8 | AY148431 Arabidops |
| 5 | 268.5 | 8.3 | 82584 | 8 | NCE13833 Neurospor |
| 6 | 266.5 | 8.9 | 15388 | 8 | ATH011977 Arabidops |
| 7 | 265.5 | 8.8 | 6688 | 8 | AF468822 Diaporthe |
| 8 | 263 | 8.7 | 5992 | 3 | DDI314910 Dictyoste |
| 9 | 262 | 8.7 | 3731 | 6 | ARI45904 Sequence |
| 10 | 262 | 8.7 | 3731 | 6 | LERDRP Y10403 L.esculentu |
| 11 | 260.5 | 8.7 | 5731 | 3 | DDI314909 Dictyoste |
| 12 | 259 | 8.6 | 165701 | 8 | AF004357 Oryza sat |
| 13 | 258 | 8.6 | 3591 | 6 | AX204840 Sequence |
| 14 | 254 | 8.4 | 3524 | 8 | AF443073 Phomopsis |
| 15 | 254 | 8.4 | 132716 | 2 | AC105931 Magnapor |
| 16 | 250 | 8.3 | 44658 | 8 | SPAC6P12 Z98533 S.pombe chr |
| 17 | 247 | 8.2 | 6847 | 8 | AY029284 Neurospor |
| 18 | 247 | 8.2 | 7178 | 8 | AF500110 Neurospor |
| 19 | 244 | 8.1 | 6987 | 8 | AY049072 Diaporthe |
| 20 | 235.5 | 7.8 | 6204 | 8 | AF411016 Neurospor |
| 21 | 230 | 7.6 | 29132 | 3 | CEFI085 Z48334 Caenorhabdi |
| 22 | 227.5 | 7.6 | 121337 | 8 | OSJN00104 AL606653 Oryza sat |
| 23 | 225.5 | 7.5 | 132699 | 8 | AC006917 Genomic s |
| 24 | 225 | 7.5 | 5025 | 3 | AF159144 Caenorhab |
| 25 | 221 | 7.3 | 4978 | 3 | AF117611 Dictyoste |
| 26 | 220.5 | 7.3 | 4013 | 8 | AF239718 Arabidops |
| 27 | 220.5 | 7.3 | 8863 | 6 | AX204839 Sequence |
| 28 | 220.5 | 7.3 | 80367 | 8 | AC012329 Arabidops |
| 29 | 220.5 | 7.3 | 104204 | 8 | ATT9C5 AL132964 Arabidops |
| 30 | 219 | 7.3 | 5208 | 3 | AF159143 Caenorhab |
| 31 | 215.5 | 7.2 | 2639 | 8 | AK101440 Oryza sat |
| 32 | 215.5 | 7.2 | 5280 | 8 | AF268093 Arabidops |
| 33 | 211.5 | 7.0 | 74406 | 8 | F2P3 AF080120 Arabidops |
| 34 | 211.5 | 7.0 | 108598 | 8 | ATT2284 AL049876 Arabidops |
| 35 | 211.5 | 7.0 | 198301 | 8 | ATCHRIV31 AL161531 Arabidops |
| 36 | 190.5 | 6.3 | 4300 | 3 | AF293414 Giardia i |
| 37 | 189.5 | 6.3 | 6103 | 3 | AF393743 Arabidops |
| 38 | 180.5 | 6.0 | 31521 | 3 | CEY47H10A AL132848 Caenorhab |
| 39 | 178 | 5.9 | 2132 | 8 | AK064209 Oryza sat |
| 40 | 177 | 5.9 | 1518 | 8 | AK064379 Oryza sat |
| 41 | 173.5 | 5.8 | 40786 | 3 | AF537961 Branchios |
| 42 | 164.5 | 5.5 | 6204 | 8 | AF411018 Neurospor |
| 43 | 161.5 | 5.4 | 4579 | 8 | PHYC11979 AJ011979 Petunia h |
| 44 | 160 | 5.3 | 38019 | 3 | CEF26A3 Z78419 Caenorhabdi |
| 45 | 151.5 | 5.0 | 80554 | 2 | AP004143 Oryza sat |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

RESULT 1

US-09-913-878A-2_COPY_710_1282 (1-573) x NCR133528 (1-4206)

QY 1 GlyValcylArgMetSerArgSerValAlaIysArgIleArgAspValLeuGlyLeuGly 20
Db 2128 GGTGTAGGCGCGATGTCGGAGACGCTGGCCAGAGAGATACGGATGTTCTCGTTGGGT 2187
QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaIysGlyMetTrpValIleAsp 40
Db 2188 GATGTGCTCTCTGTCGAAGCGCGTGTGGTTGGCCCAAGGAATGGGTATTTCGAC 2247
QY 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgIlystPglu 60
Db 2248 GTTGACACACAGCGGATGAGGATTCGATCGACATACCGTCCCGAGCGAAGTGGAA 2307
QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db 2308 TGGCACTTCGTGATTAACATCAACGTCACCTCGAAGTCGGAGCGTCTTCTGAACTG 2367
QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys 100
Db 2368 AAGTCAGCTGCTCTCAACCTTACAGCTGTTCCTGCTCTGGAAGATAGAGCCAGGACAAG 2427
QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db 2428 GTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCAACGACAGTTTCAGC 2487
QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db 2488 GAGCAAAAGCATGCTTTGAATCGCCAGTGGAAATTCGCCCAATGGGTATTTCAGAGATTAT 2547
QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db 2548 TCCAGTCGCGCAACTCGAGTCACCCAGCGCGTGTGCTTTCTTGCTGGGCTACCTGAC 2607
QY 161 SerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
Db 2608 AGTCAAGAGAGACATCGAACTTCTTGATGAACAGTGGGTTCGATCCCAAGAGACAAAG 2667
QY 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db 2668 TACTTGCAAGACATCGCTCGGATCTTCAAAAGCGGAAATGTCAACGCTTGAAGTCAAG 2727
QY 201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
Db 2728 CTGAACATCGGTGCGTGCATCAGCATACATTTACATGATGCCGATTTCTGGGTGTG 2787
QY 221 LeuGluGlnAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluSer 240
Db 2788 CTTGAGGAAATGAGTTCATGTCGATTCCTCAAAAGTTCAGAGACGAGGAGGTCT 2847
QY 241 PheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 260
Db 2848 TTTACACTCTCTATCGGACTGTGATGTCCTCGCGCGATCCCGAGCCCATTTCCCTAGT 2907
QY 261 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
Db 2908 GATATCAAGCGGTTCCAGCGATCTTCAGCCAGAGTCCACAGTCTCAAGGATGTAATC 2967
QY 281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyLysAspTyrAsp 300
Db 2968 ATCTTCTCTACTAAAGAGAGATGACCGCTTGCTGAAGAGCTATCTGCTGGAGACTACGAC 3027
QY 301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
Db 3028 GCGCATATGGCTGGGTCTGCTGGGATCCGAGATCGTCTGATGTTTCGTCATCGCGAA 3087
QY 321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGln 340
Db 3088 ATGCTCTGGAGCCGACCTGTCTAGGTACCTAAAGAGAGACAAAGACGCTTTCACAA 3147
QY 341 LeuMetAlaSerHisGlyThrGlySerAlaLysGlnGlnThrThrTyrAspMetIle 360
Db 3148 CTTATGCGCTCACACGCGACGGGCTCAGCGCCCAAGAGACGACTATACGATATGATC 3207

NCR133528 4206 bp DNA linear PLN 10-MAY-1999
LOCUS Neurospora crassa qde-1 gene, partial.
DEFINITION Neurospora crassa qde-1 gene, partial.
ACCESSION AJ133528
VERSION AJ133528.1 GI:4803726
KEYWORDS qde-1 gene; RNA-dependent RNA polymerase.
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
1
Cognoni, C. and Macino, G.
Gene silencing in neurospora crassa requires a protein homologous
to RNA-dependent RNA polymerase
Unpublished
2 (bases 1 to 4206)
Cognoni, C.
Direct Submission
Submitted (08-MAR-1999) Cognoni C., Dept. Biotecnologie Cellulari ed
Ematologia, Universita di Roma La Sapienza, Viale Regina Elena
324, 00161 Rome, ITALY
FEATURES
Location/Qualifiers
1..4206
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
1..4206
/gene="qde-1"
1..>4206
/genes="qde-1"
/function="gene silencing"
/note="putative"
/codon_start=1
/product="RNA-dependent RNA polymerase"
/protein_id="CAB42634.1"
/db_xref="GI:4803727"
/db_xref="GOA:O9Y7G6"
/db_xref="SPTREMBL:O9Y7G6"
translation="MNPIPRKNSPVEIINRLNNDYNLGIQCVDITLTPHREKEL
AEDEDFGRHKIYRALNFIYLRKDSLNQAEANPFIKAKASSNWVPAKADPDPL
WSKEPPRAATQGMALQTVLLEVRNFPNNPTGRTFGTLSPGSLGRPTSTNT
KRDEPANTVFADPPKRSRTGATPPHGAAPILKPPDPVNTGSKRPSLESNLQC
TKRAGKLSNDVAAAAPVPIASALDKVPTREHANRDPDTATGHRADQVDSFDTQS
GTYSGSVSACRHNOSTQSSFEAPSPQPREKPVDAVFEAGHLIESPKGRITKS
HIDQSLSSSOGETSFSTYSFSPSGGEGALPPRSNGLAREEBSARSQVQVHAP
VVAARLNPIWPFKWLHEAPLAVAEVTELPMECKVDESLGKYDPPSMSTARDV
TDWKLYLRDAFRGPFEPKEPNDFVTAMTGNFESKSAVALVSAVLDYDPASPTA
PLVLKPLMFEGGCLTRFRGPFPEILIPSPSTSPSPVPVPPQPGVAZEVIQ
WLTMGQSLVGRQWRAFFAKADQYKRLPEEFQLRADDPKPIIKERVHFPAETGIFTR
DVFKTSVVPAPPEVQRTFQVQSLMDLQDNNWQHLKFSRIQLGLSKTYAI
MTLEPQIRHKTLLPSGTGEVMDGVGRMSRVAKRI RDVLGLGVDVPAVGQRFQ
SAGKMWIDVDDTGDDEWLETVPQSRKWECDVDFKHORTLEVRVASLEKAGLNLQ
LPVLEBRADKVMRQAIQDRINDLQROFSEQKALNRPVEFRQWVYESYSSRATRV
SHGRVFLAGLPDSQETNLFNMGSGDFPKKQKYLQDIAWDLOKRCCTLSKLNIRV
GRSAYITMADFWGLNEVHVHVSFSPDEEFTLLSDCDVLVARSPPAHFSDIQ
RVRAVPKELSLKVDYIFSTKGDVFLAKKLSGGDYGDMWVWDVETVDGFVNAEM
PLEPDLKRLKDDTFFKQIMASHGTGSAKEQTIVDMTQKSFHALQNFLEMGNTY
KERLCYINNSVQKPAIILSSVLNVDQSKQIVFNEASWAQRLREGLIGALSPLDP
MYKSDSLGRGEPHTIIDYIKTSIARPAIDKLEAPHNAKAAKOTEDGAHFDPLA
SYTFPKEISDKSRSAALIFTLKRIEVEKEYEGLVKNKEMRSDKQPPYVVRVQVY
EKWCALTPREAWIKQNTSRPQSGKVIIRLLELFLADRENMTWALLREASTAPLYVHKSPK
VWQAGRCQALYIKQNTSRPQSGKVIIRLLELFLADRENMTWALLREASTAPLYVHKSPK
DPEVVEYLGDDDFDGIPTGNGDY"

Alignment Scores:
Pred. No.: 2,16e-246 Length: 4206
Score: 3010.00 Matches: 573
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

```

361  GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrIlys 380
3208 CAGAAGAGCTTCATTCGCGCCCTGAGCCCACTTCCTGGGCATGTGCACCTAACAA 3267
381  GluArgLeuGlyTyrIleAsnAsnSerValSerAsnLysProAlaIleIleuSerSer 400
3268 GAAAGGCTCTGTATCATCAACAATAGTGTCTAACAGCGCGGCATCATCTTAGTTCA 3327
401  LeuValGlnVasLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrp 420
3328 CTGGTGGGAACCTCGTCATCAGAGCAAGCAAGGTATGTCTTAAAGCAAGCTGG 3387
421  AlaGlnLeuArgGluLeuLeuGlyAlaLeuSerLeuProAsnProMetTyrIlys 440
3388 GCTCAATTCGATAGGAACCTGCTGGCGTGCATTCCTTCCCTGAGCCCAATGACAA 3447
441  SerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSer 460
3448 AGCGACAGTTGGCTCGGCGCGGAGAGCTACCCACATTTAGTACTCTGAAATTCCTC 3507
461  IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAla 480
3508 ATCCGAGGCTGCGATTGACAGGAACCTGGAAGCTTCCCAATGCCATGGAAGCGCC 3567
481  LysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrThrPhe 500
3568 AAGATACAGACAGCGCGCTCATTTTGGGATCCGATCTCGCTTCTTCTACTACGTTT 3627
501  PheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsn 520
3628 TTCAGAGGATTAAGCAGCAAGTCGCATCGTCCGACCTCTATTACGACTCTGAAGAAC 3687
521  ArgIleGlyValGluLysGlyTyrGlyArgLeuValLysAsnLysGluMetArgAsp 540
3688 CGTATCGGGAAGTCAGAAAGAAATATGCGAGGTTGTCGTAAGCAAGGAGATGAGAC 3747
541  SerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThr 560
3748 AGCAAGACCCCTACCTGTCGCGTCAACAGGTTTATGAAAAATGGTGGCCCATCAG 3807
561  ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSer 573
3808 CCTGAGCGGATGACAAATCCGAGCAAAATATGATTCT 3846

RESULT 2
AX034425 LOCUS 8045 bp DNA linear PAT 22-SEP-2000
DEFINITION Sequence 1 from Patent WO0050581.
ACCESSION AX034425
VERSION AX034425.1 GI:10303130
KEYWORDS Neurospora crassa
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

REFERENCE
1 Cogoni, C. and Macino, G.
Isolation and characterization of a n. Crassa silencing gene and
uses thereof
JOURNAL Patent: WO 0050581-A 1 31-AUG-2000;
UNIV ROMA (IT); COGONI CARLO (IT); MACINO GIUSEPPE (IT)
FEATURES
source location/Qualifiers
1. 8045
/organism="Neurospora crassa"
/mol_type="unassigned DNA"
/db_xref="taxon:5141"
2447. 6655
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC10121.1"
/db_xref="GI:10303131"
/db_xref="RENTREMBL:CAC10121"

CDS

```

```

/translation="KNPITPRKNSPVEIINLNNDYNGIQCVAADTLTPHRRKEL
ASDEDFGRHDKIYRALNIVLRKDDSLNQAEANFFIEAKAASNSWPKAHADPDPLP
WKEPPRAATAGQWALQTVLLEVNLRMPNNPTGRTFGRSLSGSLSPSTNT
KREDEPANTFADPKRSATGATPVHGAALPIKFPDPVNTGSKRPSLESLNQC
TKRAGKLDNVAAPVPIASALDKVTPRHRANTDPTATGHRRAQDQVSDTQ
GTSYSSFSACRHNQSTTQSSFPAPPSPREKRPVDATVFEAGHLIESFGSKRTKS
HIDNQLSSSOGETSFSTYSSFPSSGGEAIPEPSRNGLARSESARSQVTHAP
VVAERLNIWPKFPEKWLHEAPLAVAEVTRLFMHCKVDLEDESLGLKYPSPSTADVA
TDIWKTLRDLA FRGKPFPEKPNVDTMTGNPESKGSASVLSAVLDYNPNSPTA
PILVYLKPLMEQGLRRFRGDRFFELLIPSTSTSFSPVVPVSKORGAEEVIO
WLTMGHSLVGRQWRAFFARQADPRFQLRAEDPRPILIKERVHFAFTAGITTRP
DVFGRSVPAEPEVEQRTSFVSKQMLDLQDNNITWPHLFRIOLGUSKTTVAI
MTLEPHQITVHDKTLLPSGTGEVMDGVGRMSRVAKIRDLVGLSDVPSVAGVRFQ
SAGMVIDVDDTQDEDTETVPSQRKWCEDFVKHQRTLEVRVASSELKASGJNLQ
LPLVEDARQKWKROAIGDLINDIQRQFSEOKHALNPVEPROWYVYSSRATRV
SHRVPFAGLPSQETELNPLMNSCFDQKKOYLQDLDKQKCDTLKSKLNRV
GRSAYIMADFMVGLDEEVEHVFSSKFRDESEFTLLSDCYVARSAPSPDIO
RVEAFKPLSHLKDVIIFSTGDVPLAKLSDGDDYDMWVWVQFQEPVDFVGNAM
PLEPDLISRLKKDKTTFKQIMASHGTSAAKEQTTDMTQKSFHFAQLNFGMCNTY
KEBLCYVNNVSNSKPAIILSSLVNLDQSKQIVFNESWAQLELGLGALSPLDP
MYKSDMLGRGEPETHIIDLKFSIARPAIDKELEAFHNAKAAKDETDGAFWDPDLA
SVYTPKEISDKRSALLFTTLKRIEVEKEVGRVKNKEMRDKSDPYVYHNSPVY
EKWCATPEMDKSGANYDKVIRLELSFLADRENTWALLRASFATKLYYHKSPKF
VWQAGRLQIYIKQMTSRPGEAPALMTAFMYAGLMPDKKFTTKQYVARLEGDSVEP
DPEVYEVLDGDDDFDGGFTGNGDY"

```

ORIGIN

Alignment Scores:
 Pred. No.: 5,09e-246 Length: 8045
 Score: 3010.00 Matches: 573
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-09-913-878A-2_COPY_710_1282 (1-573) x AX034425 (1-8045)

QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
 DB 4574 GGTGTAGCGCGCATGTCGGAAGCGTGGCCCAAGAGGATACCGCATGTCTCGGTTTGGGT 4633
 QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
 DB 4634 GATGTGCTCTGCTGTGCAAGGCGGTTTGGTTCGCGCAAGGGAATGTGGTTATCGAC 4693
 QY 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
 DB 4694 GTTCAGCACACAGCGCATGAGGATGGATCGAGACATACCCGTCGCCAGCAAGTGGGAA 4753
 QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
 DB 4754 TCGGACTTCGTTGATTAACATCAACGATCCCTCGAAGTCGGAGCGTGGCTTCTGAACTG 4813
 QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
 DB 4814 AAGTCAGCTGGTCTCAACCTACAGCTGTACCTGCTCTGGAAGATAGAGCCAGGCAAG 4873
 QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsnAspLeuGlnArgPheSer 120
 DB 4874 GTGAAGATGCGCCAGCAATCGGTGACCGTCTTATCAACGATTTGCAACGACATTTAGC 4933
 QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
 DB 4934 GAGCAAAAGCATGCTTTGAATCGCCAGTGGAAATTCGCCAATGGTTTACGAGATTAT 4993
 QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
 DB 4994 TCCAGTCGCGCAACTCGAGTCAGCCAGCGCGTGTGCCCTTTCTTCGTGGGCTACCTGAC 5053
 QY 161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
 DB 5054 AGTCAGAGGAGGACACTGAACCTCTTGTGATGAACAGTGGTTCGATCCCAAGAGCAAAAG 5113

QY 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerIys 200
 DB 5114 TACTTGCAGACATCCGCTGGGATCTTCAAAAGCGGAAATGTGACACGTTGAAGTCCAAG 5173
 QY 201 LeuAsnIleArgValGlyArgSerAlaTrpIleTyrMetIleAlaAspPheTrpGlyVal 220
 DB 5174 CTGAACATCCGCTGGTGGATCAGATACATTTACATGATTGCGGATTTCTGGGCTGTG 5233
 QY 221 LeuGluGlnAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSer 240
 DB 5234 CTTGAGGAAATCAGGTTTCATCGGATTCCTCAAAAGTTCAGAGACGAGGAGGAGTCT 5293
 QY 241 PheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 260
 DB 5294 TTTACATCTCTATCGGACTGTGATGCTCTGCTGGCGCATCCCAAGCCCAITTCCTAGT 5353
 QY 261 AspIleGlnArgValArgAlaValPheLysProGluLeuLeuHisSerLeuLysAspValIle 280
 DB 5354 GATATCCAAAGCGGTTTCAGACAGTCTTCAAGCCAGAGCTCCACAGTCTCAAGGATGTAAATC 5413
 QY 281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
 DB 5414 ATCTTCTCTACTAAGGAGATGTACGCTTGTCTAGAGAGCTATCTGGTGGAGACTACGAC 5473
 QY 301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
 DB 5474 GCGCATATGCGCTGGGCTGCTGGGATCCGAGATCGTGCATGTTCTGCTCAATCGGAA 5533
 QY 321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrPheLysGln 340
 DB 5534 ATGCTCTCTGGAGCCGACCTGCTAGGTACCTTAAGAGGACAAACGACITTCAAACAA 5593
 QY 341 LeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrTyrAspMetIle 360
 DB 5594 CTTATGCGCTCAACAGCGACGGCTCAGCGGCGCAAGAGCAGACTACATACGATATGATC 5653
 QY 361 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
 DB 5654 CAGAAAGCTTCATTTCCCTCGACCGCAACTCTCTGGGATGTGACTAACTAACAA 5713
 QY 381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleLeuLeuSerSer 400
 DB 5714 GAAAGCTCTGTACATCAACATAGTGTCTTAACAGCGCGCCATCATCTTATTGTCA 5773
 QY 401 LeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAspTrp 420
 DB 5774 CTGGTGGAAACCTCGCTCGATCAGACAGCAAGGATTTGCTTTAAAGCAAGCAAGCTGG 5833
 QY 421 AlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLys 440
 DB 5834 GCTCAATTGCTAGGGAATGCTTGGCGGTGCATTTGCTCTTCCAGCCCAATGTACAAG 5893
 QY 441 SerAspSerTrpLeuGlyArgGlyGluProThrHisIleLeuAspTyrLeuLysPheSer 460
 DB 5894 AGCGACAGTTGGCTCGGCGCGAGAGCCCTACCCACATTTAGCTACCTGAAATTTCTCC 5953
 QY 461 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaIle 480
 DB 5954 ATCGCCAGGCTCGATTGCAAGGAACCTGGAAGCCTTCCAAATGCCATGAAGCGGCC 6013
 QY 481 LysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrThrPhe 500
 DB 6014 AAGGATACAGACAGCGCGCTCACTTTGGATCCGGATCTCGCTTCTTACTACACGATTC 6073
 QY 501 PheLysGluLeuSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsn 520
 DB 6074 TTCAGGAGATTACGCAAGTCCGATCCGCTCGCTGCTGCTATTCAGCTCTGGAAGAAC 6133
 QY 521 ArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp 540
 DB 6134 CGTATCGCGAAGTTCGAGAAAGATATGGCAGGTTGCTCAAAAACAAAGGAGATGAGAC 6193
 QY 541 SerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThr 560

DB 6194 AGCAGGACCCCTACCTGCTCCGCGTCAACGAGTTTATGAAATATGGTGGCCATCAG 6253
 QY 561 ProGluAlaValAspLysSerGlyAlaAsnTyrAspSer 573
 DB 6254 CCTGAGCGATGACAAATCCGAGCAATATGATTTCT 6292
 RESULT 3
 NTA011576 3505 bp mRNA linear PLN 25-FEB-2003
 LOCUS Nicotiana tabacum mRNA for RNA-directed RNA polymerase.
 DEFINITION AJ011576
 ACCESSION AJ011576 GI:4138281
 VERSION drp gene; RNA-directed RNA polymerase.
 KEYWORDS Nicotiana tabacum (common tobacco)
 SOURCE Nicotiana tabacum
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottepeich, F., Saenger, H. L. and Wassenecker, M. Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Tomato
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3505)
 REFERENCE Wassenecker, M.
 AUTHORS Direct Submission
 TITLE Submitted (16-OCT-1998) Wassenecker M., Max-Planck-Institut of Biochemistry Virolidresearch, Am Klopferspitz 18A, 82152 Martinsried, GERMANY
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..3505
 /organism="Nicotiana tabacum"
 /mol_type="mRNA"
 /cultivar="Petit Havana SRI"
 /db_xref="taxon:4097"
 /tissue_type="leaves"
 gene 1..3505
 /gene="RDRP"
 CDS 20..3370
 /gene="RDRP"
 /codon_start=1
 /evidence=experimental
 /product="RNA-directed RNA polymerase"
 /protein_id="CAA09697.1"
 /db_xref="GI:4138282"
 /db_xref="GOA:Q9ZRY7"
 /db_xref="SPTREMBL:Q9ZRY7"
 /translation="MGKTIQVFGFPYLLSAAVKLEHNTGNGTVYALVSKSGKGR
 RAFAKVQFANKSAEFTIDASKGLYSPYLSKAWEMKTDIVQPTTYVHMDGVTTLNF
 GCOISDKKAVLGSKDVSITKFGIGLAKIVFPLSHADYKQLSYENIQWVLRHYG
 QNAQFILLQLFGAPRIYKLESDSCYSPFKETPDQWRTTDTFTPSVGLSSCLBFR
 NGVLPNFSSEFFYIKESNNQFLOQFTFSSQKALVPIVHLPSGILPYKILFKI
 SSIHQGCPFLALNFFQVLDVRRNRNACIEHALEKLYLKECCIDPVRWLTQYD
 EYLKQQLKSPPTITLDDGLVYVRVVTVCYVCGPVNVSRLNRYSEIDNNFL
 RVSGVBEWEKIHSTDLLPRASTNGTRTDIYERILSTLRNGFIIDGKRFELAFSS
 QLRDNSWFMASRPLTANDIRTWGDFRIRNVAKYAARLQSGFSRRTYSGRHE
 VEVIPTDVACSLHGTNIIFSDGIGKIADFAHRVTIKGQYTPSFQIRYGGYGVVA
 VDPSSKLSLRKSMLYKSSNNIKLDVLGSKYQPCYLNRLQVTLTLSTLGIIDVLEQ
 KQEAVDLLDRLHSLKAGEALELMSPEENTILKELNCGYMPDAEPFLSNLOTF
 RASKLLDRLTRTIFIPNGRTMMGCDSESTLEYGVTFVQFSAGRQRPEESHEND
 SGSAACDFILGNVVAKNPCLHPGDIRLVLRADVLPALHNVDCVVPFQKGRPHNE
 CSGSDLDGDIYFCVWDPIPPRQSGMDYTPAFTQLDHDVTIEVEEYFTNIIND
 SLGI1ANAHVVFADRPDMASDPCKLAQSFIAVDFFPKTPGPAITPSQLRPEYDI
 FMKPKDPTPTSEKVIKGLQKVNKTTPQASITATPRDVARKSYSDMTVDGPEYI
 DEAPYKSEYDNKLGMDYVGIKTEAEILSGGIMKASKTFDRKDAEALGVAVRCLR
 KEARAFKRSDIDMLAKASAKVHYVTHHTYGLYNEGLKRDHFIFFPCWCVVDQLIQ
 IKKAKARKRPVPHLSLSLSLSKLVIN"

ORIGIN
 Alignment Scores: 1.7e-13 Length: 3505
 Pred. No.:

| | | | |
|---|--------|--|-----|
| Score: | 275.50 | Matches: | 123 |
| Percent Similarity: | 35.65% | Conservative: | 82 |
| Best Local Similarity: | 21.39% | Mismatches: | 215 |
| Query Match: | 9.15% | Indels: | 155 |
| DB: | 8 | Gaps: | 16 |
| US-09-913-878a-2_COPY_710_1282 (1-573) x NTA011576 (1-3505) | | | |
| QY | 1 | GlyValGlyArgMetSerArgSerValAlaLysArgGlyAspValLeuGlyLeuGly 20 | |
| DB | 1604 | GGAATTGGAAAAATATCTGCTGACCTTGCATAGAGTTACCAATAAATGGCCCTTCAA 1663 | |
| QY | 21 | AspValProSerAlaValGlnGlyPheGlySerAlaLysGlyMetTrpValLeuAsp 40 | |
| DB | 1664 | TATATCCATCTCTTTTCAGATTGCTTATGCTGATATAAAGTGTGTGCTGTTGAT 1723 | |
| QY | 41 | ValAspThrGlyAspGluAspTrpIleGluThrTyProSerGlnArgLysTrpGlu 60 | |
| DB | 1724 | -----CATATTCAATCAATGAAG----- 1741 | |
| QY | 61 | CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80 | |
| DB | 1742 | -----TTATCTTTGAGAAAGAGCATGTTGAATAT 1771 | |
| QY | 81 | LysSerAlaGlyLeuAsnLeuGlnLeu-----ProValLeuGlu 94 | |
| DB | 1772 | GATCAACAAACATAAGATTAGATCTCTGGGATGGAGCAAAATATCAGCCTTGTTATCTT 1831 | |
| QY | 95 | AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 114 | |
| DB | 1832 | AATCGTCAACTGTTACTCTCTGTCTACTCTTGGGTGATAGATGATGTTTGTAGAGCAG 1891 | |
| QY | 115 | LeuGlnArgGlnPheSerGlnGlnLysHisAlaLeuAsnArgProValGluPheArgGln 134 | |
| DB | 1892 | AAGCAAAATGAGCAGTAGATGATGCTGATGCTATTTGTCATGATCTTTTGAAGGCACAG 1951 | |
| QY | 135 | TrpValTyGlnSerTy-SerSerArgAlaThrArgValSerHisGlyArgValProPhe 154 | |
| DB | 1952 | GAAGCTTTAGATTGATGCT----- 1972 | |
| QY | 155 | LeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPhe 174 | |
| DB | 1973 | -----CCTGGAGAGAAATCTAATATCTCAAGAGAGATGCTAAATTTGGTTAT 2020 | |
| QY | 175 | AspProLysLysGlnLysTyLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 194 | |
| DB | 2021 | ATGCTGATGCTGAGCCGTTCTTCAATGATGCTGCAACTTTCGGGCTTCCGAAGTTG 2080 | |
| QY | 195 | AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyIleTyMetIle 214 | |
| DB | 2081 | CTCGATTTCGGAGCTAGAACAAATATTTATTCCAAATGGTAGAACCATGATGGGATGC 2140 | |
| QY | 215 | AlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer----- 232 | |
| DB | 2141 | TTGGATGAATCCAGAACCTGGAAATATGTCAGGTGTTGTTTCAGTTTTCTGGTGCTGGA 2200 | |
| QY | 233 | -----LysPheArgAspGluGluSerPhe-----ThrLeuLysSerAspCysAsp 248 | |
| DB | 2201 | CGTAGACAGTTTTTTGAGGAATCGCATCTCTTTAATGACATGGATCCGCCAATCTGTAT 2260 | |
| QY | 249 | -----ValLeuValAlaArgSerProAlaHisPheProSerAspIle 262 | |
| DB | 2261 | TTTCATCTCAAGGAAAGCTGGTCTGTGAAAAATTCATCCTGTCATCTCCGGTGATATT 2320 | |
| QY | 263 | GlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIlePhe 282 | |
| DB | 2321 | CGTGTTTTAAAGGCTGTTGATGTTCCAGCTTGACCAATGGTTGATGTTGTTGATTT 2380 | |
| QY | 283 | SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyAspGlyAsp 302 | |
| DB | 2381 | CCCCAGAAAGGAAAGAACCTCATCCAAATGAATGTTCTGGAAGTGATTTGGATGGGAT 2440 | |
| QY | 303 | MetaLtpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322 | |

| | | | | |
|------------|---|--|------|------------------------|
| Db | 2441 | ATCTACTTTGTTGTTGGGATCCAGACTTGATCCGCCAGGCAAGTCCAGTCGATCGAT | 2500 | |
| Qy | 323 | LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMet | 342 | |
| Db | 2501 | TATACTCCAGCACCCACACACAGTGGATCATGAT----- | 2536 | |
| Qy | 343 | AlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyAspMetIleGlnLys | 362 | |
| Db | 2537 | -----GTCCACATTGAGGAAGTTGAAGAG | 2560 | |
| Qy | 363 | SerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyLysGluArg | 382 | |
| Db | 2561 | TACTTC-----ACCAACTAT----- | 2575 | |
| Qy | 383 | LeuCysTyxIleAsnAsnSerValSer----- | 391 | |
| Db | 2576 | -----ATCATTTATGACATTTGGGAATCATAGCAAAATGCGCACGTTGTTATTGCGAGC | 2629 | |
| Qy | 392 | -----AsnLysProAlaIleLeuSerSerLeuValGlyAsn | 404 | |
| Db | 2630 | AGAGAACCTGATATGGCCATGACGATCCGTGCAAAACAACCTGCTCAGTCTTTTCAATT | 2689 | |
| Qy | 405 | LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg | 424 | |
| Db | 2690 | GCAGTTGACTTTCCTCAAGACTGCGGTTCCAGCTGAAATACCATCT-----CAGTTGCGC | 2743 | |
| Qy | 425 | ArgGluLeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyLysSerAspSer | 443 | |
| Db | 2744 | CCTAAAGAAATACCCAGACTTCATGGAAGAGCCAGACCAACCTATCCCTCGGAAAAA | 2803 | |
| Qy | 444 | TrpLeuGlyArgGlyGluProThrHisIleLeuAspTyLysPheSerIleAlaArg | 463 | |
| Db | 2804 | GTATTGGAAAG----- | 2815 | |
| Qy | 464 | ProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThr | 483 | |
| Db | 2816 | -----CTTTCCAGAAAGTGAAGAACAAA | 2839 | |
| Qy | 484 | GluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyTyThrPhePheLysGlu | 503 | |
| Db | 2840 | -----ACACCCAGGCTAGCTCCATCGCGACCTTCACAGGGAT | 2878 | |
| Qy | 504 | IleSerAspLysSerArgSerAlaLeuPheThrThrLeuLysAsnArgIleGly | 523 | |
| Db | 2879 | GTGCGCAGGAATCATATGATAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG | 2938 | |
| Qy | 524 | GluVal-----GluLysGluTyGlyArgLeuValLysAsn | 535 | |
| Db | 2939 | GAAGCTTTTACTACAAAAGTGAATACGACAAACAGCTTGGTAAT | 2983 | |
| RESULT 4 | | | | |
| LOCUS | AY148431 | 3807 bp | mRNA | linear PLN 12-NOV-2002 |
| DEFINITION | Arabidopsis thaliana RNA-dependent RNA polymerase 1 mRNA, complete cds. | | | |
| ACCESSION | AY148431 | | | |
| VERSION | AY148431.1 | GI:24935274 | | |
| KEYWORDS | | | | |
| SOURCE | Arabidopsis thaliana (thale cress) | | | |
| ORGANISM | Arabidopsis thaliana | | | |
| REFERENCE | 1 (bases 1 to 3807) | | | |
| AUTHORS | Yu, D., Fan, B., MacFarlane, S.A. and Chen, Z. | | | |
| TITLE | Analysis of the involvement of an inducible Arabidopsis RNA-dependent RNA polymerase in antiviral defense | | | |
| JOURNAL | Unpublished | | | |
| REFERENCE | 2 (bases 1 to 3807) | | | |
| AUTHORS | Yu, D., Fan, B., MacFarlane, S.A. and Chen, Z. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (05-SEP-2002) Botany and Plant Pathology, Purdue | | | |

```

FEATURES
Source      University, West Lafayette, IN 47907, USA
Location/Qualifiers
1..3807
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
128..3451
CDS
128..3451
/notes="RDEPI"
/codon_start=1
/product="RNA-dependent RNA polymerase 1"
/protein_id="AA064409.1"
/db_xref="GI:24935275"
/translation="MGKTIQVGFNGVSAEYKFKFLERLTGSGTVVAIKVRQPKGKG
PRVYATVQTSERHRLIITAAERLYYRSYLKAFVEYQDIPKPRASLHTISLKM
FGQCVSTKFKLTLSAODVCYSGIKGWEKLFHSPSKVOKYVLELSVENTWQIDLH
POGRSKFLVIVIGAPKIFEXEDQPINLLFGIMDFYSDGDEQWIRTSCTSIG
QSTACLEPFLNVPDFRNPANTAEHRASSFLLESSSYSSNANTUVPVDDPPGF
SUPPELFKNTLVONACLSGALDLYRLLNKKYDRALIDHCLKLFLHGCCVE
PARWRDEYKWIISGKPLISPTISLDLGLVYRVQVTPFARVFSGEVNVNVRVL
HYSKVINPLRYSFVDEDLKVRISMDLSPRSQRTKLVDRIYSVLRDGIVIGDKPF
EFLATSSQLRENSAMWAPIDRIITAAHIRAMWGDPIHNVAKYARLGGSPSSRE
TLNVRSDTEVTPDVELISLGRVYVFSGLGKISAEFARKVARKGLTFEFSFAFQIR
YGYKGVAVDPSNKKSLRSKMSKPSSENTKLDVLANSKYQPCYMNRLIITLSL
GVDSVFKEQREVDRDLALHLEAHEALGLMAPGENTNLIKALICGKPDAP
FLSMQNPFRASKLELRITKTRIFISGGRSMGCLDETRELEYGVVVQYSDPMRGR
REIITGVVVAKNPCILHSGDVRVLOAVNPALNMVDCVFPQGLRHPHNECSGDL
DGIYFVCDQLPPTSEPMYTPPTQLIDHDVTIEEVEYFANTYVNDISGIIA
NAHTAFADKEPDLSPDCIELAKKFTAVDPKGTGVAAVIPOHLVYVKEYPDMKEPD
KPTYESKNVIGLFEVKEPAPLISKFTLDVASKVDKMDVDGFEVVDFAFYQ
KANYDFKLNMDYGIKTEAIIISGIMMSKSFTRKRDASIGRAVRLAKSTLSL
FNASBEENESAKAWHTVTHSSYIWLNGEHLNRDHLFSFAMCVKDKLVRIKNTLN
GRROQETLERLDHLRFQ"
ORIGIN
Alignment Scores:
Pred. No.:      7,47e-13      Length:      3807
Score:          268.50      Matches:     137
Percent Similarity: 37.90%      Conservative: 90
Best Local Similarity: 22.87%      Mismatches:  225
Query Match:     8.92%      Indels:      148
DB:              8          Gaps:         22

US-09-913-878a-2_COPY_710_1282 (1-573) x AY148431 (1-3807)
QY      1 GlyValGlyArgMetSerArgSerValAlaLysArgAlaArgAspValLeuGlyLeuGly 20
DB      1736 GGAATAGGAAGATATACGCTCAATTTGCTAGAAAGTAGGAGGAAATGGTCTTACA 1795
QY      21 AspVal---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMet-TripVal11 39
DB      1796 GAGTTTTCCTCATCTGCTTTTCAATCCGTATAGCGGCTTATAAGAGAGTGGTGGCTGT 1855
QY      39 eAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTr 59
DB      1856 G----- 1856
QY      59 pGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgServAlaLaserG1 79
DB      1857 -----ATCCAACTCATCAAGAAA-CTGCTCTGAGGAAGAGATGAGGAA 1902
QY      79 uLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeu-----ProValle 93
DB      1903 ATTGCAATCGGAACACCAAGCTTGATGTTCTGCGGTGGAGCAAGTACCAACCTGTGA 1962
QY      93 uGluAspArgAlaArgAspLysValLysMetArgGlnAlaLleGlyAspArgLeuLeAs 113
DB      1963 TATGACAGCAACTGATTACTTTTGTCTACTCTTGAGGATTACGACAGTGTGTTGA 2022
QY      113 nAspLeuGlnArgGlnPheSerGluGlnLysHisAla---LeuAsnArgProValGluPh 132
DB      2023 GAAGAAACAAAGGAAGTTGTGATCGTCTCGACGCAATCTTGACTCATCTCTTGGAA-- 2080
QY      132 eArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgVa 152

```

```

2081 -----GCTCACAGGCTCT 2094
QY      152 lProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSe 172
DB      2095 TGGTTTAATGCT-----CCAGGGGAACAACAATAATTCACAGGCATTGATCTGTG 2148
QY      172 rGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysAr 192
DB      2149 TGGCTATAAACCTTGACGCTGAACCTTTCCTTCAATGATGCTTCAGAAATTCAGGCGATC 2208
QY      192 gLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaLysIleTy 212
DB      2209 TAAGTTGTTGGAATACGGACAAAACAAGGATTTTCTTCTGGTGAAGATCTATGAT 2268
QY      212 rMetIleAlaAspPheTrpGlyValLeuGlnGluAsnGluValHisValGlyPheSerSe 232
DB      2269 GGGATGCTTAGACGACGACCAAGAACGCTGGAGTAGTGTTCAGGTAGTAGTCAGATTAATC 2328
QY      232 rLysPheArgAspGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 252
DB      2329 TCCATGAGGCCCGGGAAGGGA-----TTCATCATCACCGGACCTGTGTTGTTGTC 2379
QY      252 aArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGl 272
DB      2380 CAAAAACCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2439
QY      272 uLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLy 292
DB      2440 TTTAATCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2499
QY      292 sLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu11 312
DB      2500 TGAATGTTCTGGAGTGATTTAGATGGAGATATTTACTTTGTTGTTGTTGTTGTTGTTGTT 2559
QY      312 eValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLys 332
DB      2560 GGTTCGCCCAAGAACGCTGCAACCAATGACATACCTCTGCAACCACTCAATCTTGA 2619
QY      332 sLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLy 352
DB      2620 TCATGAT----- 2626
QY      352 sGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPh 372
DB      2627 -----GTCACAATTTGAGGAAGTTGAGAGTACTTTGCGAATCACTATTGTAATGATAG 2679
QY      372 eLeuGlyMetCysThrAsnTyr-----LysGluArgLeuCysTyr11 386
DB      2680 TTTAGGATCATCGCAATGCTCATACCGCTTTCGCGATAGGAACCACTC----- 2731
QY      386 eAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVa 406
DB      2732 -----AAGCGCTTTAGTGACCCATGCATTTGAGCTTGCAGAAAGAGTTTTCACCTCCGT 2784
QY      406 lAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg1 426
DB      2785 AGACTTTCMAAAACTGGTGTGCA-----GCTGTGATACCGCAACA 2826
QY      426 uLeu-----LeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSe 441
DB      2827 TCTTTACGTGAAGAGATATCTGATTTTCATGGAGAACCGGATAGCCACATATGATGTC 2886
QY      441 rAspSerTrpLeuGly-----ArgGlyGluProThrHis11 453
DB      2887 GAACACGCTATGTTAGCTTTTATAGAGGTTAAAGAGCGAGCTCCACCGTTGATCTC 2946
QY      453 eIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu----- 470
DB      2947 GATCAANTCATTCAACGCTTATGTCCTCAAAAGTCTTATGATAAAGATGGAAGTTGA 3006
QY      471 -----GluAlaPheHisAsnAlaMetLysAlaAlaLysAspTh 483

```



```

Db      3007  CGGATTGAGGAGTATGTTGATGACGCTTCTAC-----CAGAAGCGCAATTATGATT 3060
QY      483   rGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheGlyGlu 503
Db      3061  uAAGTTAGGT-----AATTGATGGATTACTACTATGGGATTAAAGACTCA 3102
QY      503   u-----IleSerAspLysSerArgSerSerAlaLeuLeuPhePhe 516
Db      3103  GGCTGAGATCTTAGTGGTGTATATGAGGATGCAAGTCA-----TTTAC 3150
QY      516   rThrLeuLysAsnArgIleGlyGlu---Val-GluLysGluTyrGlyArg----- 531
Db      3151  C-----AAGACGCGAGATCGCAATCGATTGCGAGAGCGGTTCCGGCGCTTGAGGAAGA 3204
QY      532   -----LeuValLysAsnLysGluMetArgAspSerLys 542
Db      3205  AACTTTGCTGTTCTCAATGCTTCTGAGAGAGAAATAGTACGCAAGG 3257

RESULT 5
NCBI3B3/c
LOCUS
DEFINITION Neurospora crassa DNA linkage group II BAC contig B13B3.
ACCESSION BX284762
VERSION BX284762.1 GI:28881246
KEYWORDS
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE
1 Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
Unpublished
2 (bases 1 to 82584)
German Neurospora genome project.
Direct Submission
Submitted (06-MAR-2003) MIPS, Institut fuer Bioinformatik,
GSP-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
G.Mannhaupt@gsp.de Project Coordinator: Ulrich Schulte, Institute
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
E-mail: ulrich-schulte@uni-duesseldorf.de
this contig is an assembly of BAC 13B3 from 1 to: 11950 and BAC
11C11 from 11951 to: 82584, strain OR74A, BAC clones are available
at the Fungal Genetic Stock Center http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
http://www.mwgdna.com
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of linkage groups II
and V can be viewed at: http://mips.gsf.de/proj/neurospora.

FEATURES
source
Location/Qualifiers
1..82584
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
/chromosome="6"
repeat_region
3959..3982
/note="24 bp tcatcatcatca tandem repeat"
gene
5038..6880
/gene="B13B3.010"
join(5038..5145,5216..6880)
CDS
/note="B13B3.010"
/note="similarity to clone mgc:8691, similar to amplified
in osteosarcoma, homo sapiens, TREMBL:BC000532.1"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD70506.1"
/db_xref="GI:28881247"
/translation="MRRPSSLALLSLPFGSARQPSAFSIHQDLAHPQEVFSDS
VYFEADAFALLAANKTKPTKTPASDAGKSTRTDLTSAIRESATANADTDGDESI
GCTSLRETYELIHPMPMYLCSIPLIAPPALANKTATELAKAEAEVTRVYKNGE
LMRGLEHCVGSMWYQVYCKSIYQVYHAVENPKGFPLEBKNSQILGSLPEP
SSHSQKQKQIEVPNNEQQLSPFPNTELQAKDNQRYLVQRLDGGTICDLTGPRITIEI

```

```

OYHCNPALSDRIGWIKVEVTTCAYLWVHTPRICADVAFLLPKPKTKAHPITCQIITS
DEBALSPNRRKNTIDSAASAAAATTTEDQKSESSESPKLSYGGUTVAGIPGARR
ILPSTHVLPLQROORQAQQNLEALTAAAFADVDFDYGDNNNNNNHHHPKA
KGRKAGAGKGGQGGKMKMRISERIDKGLDQOTDLALREETFAAGLDPDFNLN
DERAGEIIVVEFVADVSGDEDDGAEGAIEKEVFWYVEDEGEPAEAGKDKQESKK
GGNGEGSGSGSEBGSKEEYVDEL"
5038..5145
/gene="B13B3.010"
/number=1
5146..5215
/gene="B13B3.010"
/number=1
5216..6880
/gene="B13B3.010"
/number=2
7534..7562
/note="29 bp poly(a) tandem repeat"
12653..15824
/gene="B13B3.020"
join(12653..12719,12910..13083,13261..13438,13497..14132,
14195..15824)
/gene="B13B3.020"
/notes="similarity to transcription activator, Aspergillus
oryzae, PIR:JC7252
Contains Fungal Zn(2)-Cys(6) binuclear cluster domain
[VKCKRLRRICDYQSPQPCIDAKLQC]"
/codon_start=1
/product="related to transcription activator"
/protein_id="CAD70507.1"
/db_xref="GI:28881248"
/translation="MMFMWGHRRGRNLAIPGLQLRSSSPFAQIVLKTPEPTGS
TKQSPRPSSSSNATSAAGSNRPVQPKVKDCRKLRRICDYQSPQPCIDAKLQCTR
NHVPRKGRPFGRGLINLREREAGKEQRPSSGDPDPSDARSNNTDIYFONAS
VSSTWASDSTAPETGLMANEHI FSRGAFYLI PRCVELFYCHLYPVNVLVMPRI
QRMDPRVDLPKRNLLFSALTCFRMSGHSLGAEDSPFWDQAGFLNDCLDVRKQ
YDLYDNISLSTVSSIFLASSFPETNOSTKAWYLRALTFACELGLEDESTYAGLSP
FEALCQRVFWLLYNNERSFALIRNPLMRAPKLPWTLHAYESGHTGFLQLLGI
YIPRDSILIAWTYNSIVDVNTYALQNQLARPAGSSGGLSIPHPFPSSPNAAVS
SIQMPVSCQSFDLSLSSSSSPEDVPVVPVETASQTAELVTQWRLIILWS
SLRQGLYWAENESMHFALPLTARQTGLVGRSLNQIGLEATGVALEVMGMIFE
KIFEIGACMNVLDSDKASLEQGLTLNFRGRHGAAGKKNKAARSATSSPADQVM
KDEGDEVDLLDFMRALSATPTSRKQFAEPLVMATTPGGMRIKPGSVLHDTGGG
TGAFMDAQTPGLAEQIPGLNGOGLDHPASGIGAAGHDASVWVHQQQQOQVWNA
SAYAGCAQLPGSATSGAGSLRDPDGGGGGCGQCMEDVVMGDQASHQGGQGNIDPR
LAGSNVPRGLQHAIFRPLRNAPLPHLSPFGQFGGPTSLYDVGWSPIDPFGGLT
RAASRTMGMLRGLMAMVQGGPTGSGSPSVSQGS"
12653..12719
/gene="B13B3.020"
/number=1
12720..12909
/gene="B13B3.020"
/number=1
12910..13083
/gene="B13B3.020"
/number=2
13084..13260
/gene="B13B3.020"
/number=2
13261..13438
/gene="B13B3.020"
/number=3
13439..13496
/gene="B13B3.020"
/number=3
13497..14132
/gene="B13B3.020"
/number=4
14133..14194
/gene="B13B3.020"
/number=4
14195..15824
/gene="B13B3.020"
/number=5

```

gene 16797..17552
/gene="B13B3.030"
complement(join(16797..16985,17226..17552))
gene /gene="B13B3.030"
complement(join(16797..16985,17226..17552))
CDS /gene="B13B3.030"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD70508.1"
/db_xref="GI:28881249"
translation="MHFKSLISLGLSLRYHAALPTTSDDTINDDTTINDDTILI
PVMIPKPNMNFNNKPNVTYKISITONPDHVDTLICPSRALLSSGPPPEPGRFST
TLGRWTVSWYQNTAYTKTLAEQAMVLDQEKRTTDTATNTRPLVAGEAVYGHDAW
SVLKGVDVCT"
complement(16797..16985)
exon /gene="B13B3.030"
/number=1
complement(16986..17225)
intron /gene="B13B3.030"
/number=1
repeat_region 17029..17053
/note="25 bp poly(t) tandem repeat"
exon /gene="B13B3.030"
/number=2
gene 18460..24617
/gene="B13B3.040"
complement(join(18460..24111,24165..24617))
gene /gene="B13B3.040"
complement(join(18460..24111,24165..24617))
CDS /gene="B13B3.040"
/note="similarity to am-toxin synthetase (amt), alternaria
alternata, TREMBL:AF184074.1
Contains Putative AMP-binding domain signature
[TOFTSGSGTKPK],ATP/GTP-binding site motif A (p-loop)
[AAMSAGK],Cysteine proteases inhibitors signature
[QQVQGLSFLGSLN],Phosphopantetheine attachment site
[AVGDSLRAMELIAAA]"
/codon_start=1
/product="related to AM-toxin synthetase (AMT)"
/protein_id="CAD70509.1"
/db_xref="GI:28881250"
translation="MQQVVGALSLFLINGQSNATSLHFDLTGLTYSRPLSLPSQK
NISLDAISLALGLLQREGEDDRVENFNWRRSDGNETSQSRVAVLGSBLDQ
TASVKNVLELVKTVQDVTEENRTLFFNDGKEAPEKIEGGEYETAEQRNFKPQELL
ASEDGQHDITLQAYFNKVPDALTSSTQIRINAFTELLGLIFSPETPLISSLLAP
LPCLNQIWFANVAVPLIDRTMQDIISQAAANPKPAVSWDGEFTVAEVESKDL
LARHLSRGVKGVLGGVPLMPKSRMTIIVALAMKAGSAFALTDTSOPEGLRVRVVE
QTQDIIIVASQTELAELVCLPSSGSGOGVITVSDDLFTTLSTPBLTSLVLPVP
TQTPPLIQTSSGTPKPGVLSHANTVSGALPRAEAVGKSSRCFEFAFADFVS
IDCMCTLAVGGTICPSDADRNDLGGAILKSGANAMHTFSVARVLDPEVIKGLDV
LGIGGEVSAADAASAKSGKTSIVVAGPSECTVGVVNVFARDEAKKGFITGNI
GTGVGGCVYVDNDHDLVPVAGVGLLEMEGVVGLGYLGEETKTFVIEDPKWL
AGHDAPHPGRKGLYKTGDLVYDADGSGAFVFGKDAQVLRQORVELVEIRHLR
AKLPSPGVKIAEVKPSGGDPTLVAFLETSTGISAQDALQSVVEGVSFPELQAA
LDGEEVLGIDLPRYVNVNAYIPLREMPSLVSGKIDRKTILRAMGAAMTREQVSGAKK
ARQGGEGAGTTMEMALHKWKLGLTQEBIAAGDSFFAVGDSILRAMRLIAAAREAS
GKVLTVADVPYVFLRDMVAVTKVDITQDQAGKNNIEIAPFSLISDWAIEDVKAAS
KNCNIDKDALEDVYPTPLQEAALNALSKEAYVAQVLRQMEGQAADKLKAAEAI
NADCAILRTIIRSPGLMGVVKVLELHWSATTLAEYLETDNEMDLGRPLVRGL
IQGDAYHIVLTHALYLDGWSMPLVVDVQYGLVPRKFAAEKPHFINLNNTLL
RAACDAHWKQLAGATQFPALPPEGYQTRADSLWELVSLDGRKPPSPNTITLA
TIVPAALVASQYQAGNNDIIFEGEITGNAPIVGADIEGEMITTVVPRVQIDREQ
TVARYLQVAAEQMITQIPYEHAGLOHRLRSLDPAOACELRTGILVHPAAGEVGTEN
TPANGLYPAGSEAKALAFNTYALMLVCSLEADGFLWASDFSCTVEMDVVKKALD
QLQVYFOLCEGDGGEKVGQLECLTEKHEEVKLVAGVKNLTPAESDPPFVGVTABD
EGWIVDAADTARLVPCVAGELIVETSKEDAPVLLAETPAWYEGNKNYKTGOLA
KFKYDGSASCLILKASEPTEVAKKVASVPSVAATSAQKTLRIWAKLKV
EDSYLDGSPFNRRGDSITAMKLVSQOGLKISVAQFANETLFDMANVMQTTGV
VDASEKETEYKPLPSLEENKASIMEVVKHLSADKSWTADILPTPLQETAVRTVI
ELPREFSLRPGOVQDKAALFRACHVEARNEILRTVFDASNTQVAGVIDNLI
TSAEYFIDHDADVKDPAQVCRLLDSQTRMPYGSSTFKMPPFLTNQTSQSLAPLSHA
QYDEICLPIMLHQHQLYLNLPSSPSSIPSPFPHFVAHTLURDSIPAAIPKWRNLLSGS

SGITRPRPSTPITSKGFALHPRAVDISSRPRDVTIATLPSAAWALALAHVKSPDVL

Alignment Scores:
Pred. No.: 4,37e-11 Length: 82584
Score: 268.50 Matches: 123
Percent Similarity: 37.74% Conservative: 57
Best Local Similarity: 25.79% Mismatches: 168
Query Match: 8.92% Indels: 129
DB: 8 Gaps: 20

US-09-913-878a-2_copy_710_1282 (1-573) x NCBI13B3 (1-82584)

Qy 1 GlyValGlyArgMetSerArgSerValAlaLysArgIle---ArgAspValLeuGlyLeu 19
Db 42932 GCGTTGGGATAATTTCGACGGGTGCCCTCGAGGTGATTTCATCGCAGATCCCGAATCC 42873
Qy 20 GlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIle 39
Db 42872 AAGGGCTATCCCAATTGCTTACAGGTTCGATGGCTGGAGCAAGGCAATGTTGGCTCTC 42813
Qy 40 AspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrp 59
Db 42812 GATGCCCGGCTCACTGGAGGCAATCTGTATCAGG---GATTCATCGAGAAG--- 42762
Qy 60 GluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGlu 79
Db 42761 -----TTCGCGACCGCTGATGAGGAACATCTCGAAATATGCACATGGCTTCAAG 42711
Qy 80 LeuLysSerAlaGlyLeuAsnLeuLeuLeuProValLeuGluAspArgAlaArgAsp 99
Db 42710 CCAATTCTCTGATGCTGAACCGCTCAGATGATCAAAATTTCTCGAGGAC--- 42663
Qy 100 LysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsnAspLeuGlnArgGlnPhe 119
Db 42663 ----- 42663
Qy 120 SerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSer 139
Db 42662 -----ATCGCGCCCTGCC-----CAGTGTCTTTGGAGCTT 42630
Qy 140 TyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuPro 159
Db 42629 CAAGAAAAGGAACATCCAACTG-----CTGCGGGAATCAACG 42594
Qy 160 AspSerGlnGluGluThrLeuAsnPheLeu----- 169
Db 42593 GATAACGTCAGAACGTCGCAACCTTTCTCAAGCTGCAGTGTGTCGGAGATCTCTGTACAC 42534
Qy 170 -----MetAsnSerGlyPheAspProLysLysGln 179
Db 42533 CTCAGCCAGTTTCTCAAGACCTTGACAAAATGAAC-----ATCGACTACCGGAGAC 42480
Qy 180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSer 199
Db 42479 CAGTTTCTTCGCGTATTGTCGAAGCTGCTGCTTAGAGCTGCGCTTGTCTCAAGCAC 42420
Qy 200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
Db 42419 AAGGCTAGATCCCGGTTCCCTATGGGCTTACTCTGTTTGGTGTGTCATGGAGCAGACGCG 42360
Qy 220 ValLeuGluGluAsnGluValHisValGlyPheSerSer-----LysPheArgAsp 236
Db 42359 CTATTGCGCGAAGCGAAGTAGTACGTACTTTGAGACCGGCGGTCGCTTCAAGCAC 42300
Qy 237 GluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAla 256
Db 42299 CTTCCCATCTGCT-----GGCCCCGTTGTTGGTGTGATCGACGAGCTCCAGCC 42258
Qy 257 HisPheProSerAspIleGlnArgValArgAlaValPheLysPro-----GluLeu 273
Db 42257 TTGCACCTGGAGACATCCAGATTGCCACAATGCCATCCACCGGAGGACCCCTCTC 42198
Qy 274 HisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAlaLysLys 293

```

Db 42197 AGAGAACTGAAACATCGATCGTGTTCAGTCAAAACGGTGAACGATGACCTCCCTAGTCAG 42138
Qy 294 LeuSerGlyGlyAspTyAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
Db 42137 CTGTCGGGCGGTGACTTGTGATGGCGACACTTTCACAGTGATCGTGATCAAGCATCGTT 42078
Qy 314 -----AspGlyPheValAsnAlaGluMetPro-----LeuGluPro---AspLeuSer 328
Db 42077 GCCAGCTCAGAACTTTCGGCGCGCCGATATCCAGTCGCGAGCCCTTGAACACTGAAC 42018
Qy 329 ArgTyrLeuLys---LysAspLysThrThrPhe-----LysGlnLeuMetAlaSerHis 345
Db 42017 CCGAGGTCGATCAAGACATGCTGACTTCTTCGTGGAGTTTATGAAGCCGATCAC 41958
Qy 346 GlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHis 365
Db 41957 CTGGGGGTAAATTCGGGTTCGGCACATGATCCCTCGCGGAC----- 41919
Qy 366 PheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyr 385
Db 41918 -----GAAAGC----- 41913
Qy 386 IleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeu 405
Db 41912 ---AATGAGGAAACGCTGGATGCAGACTGTCTCAAGCTGGCAGCATTCGACTCTAAAGCG 41856
Qy 406 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg 425
Db 41855 GTTGACTTCTCAATCGGGAT-----CACGTGGATATCACCAGATTACCAAGG 41805
Qy 426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp 442
Db 41804 -----CCACCAATGTATCGGCCTGAT 41784

RESULT 6
ATH011977
LOCUS Arabidopsis thaliana partial RDRP gene. 1538 bp DNA linear PLN 12-MAR-2003.
DEFINITION Arabidopsis thaliana partial RDRP gene.
ACCESSION AJ011977
VERSION AJ011977.1 GI:4127462
KEYWORDS RDRP gene; RNA-directed RNA polymerase.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1
AUTHORS Schiebel, W., Pellissier, T., Riedel, L., Thalmair, S., Schiebel, R.,
Kempe, D., Lottspeich, F., Saenger, H., and Wassenecker, M.
Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from Tomato
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1538)
AUTHORS Wassenecker, M.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1998) Wassenecker, M., Max-Planck-Institut of
Biochemistry, Viroidresearch, Am Klopferspitz 18A, 82152
Martinsried, GERMANY
FEATURES
Location/Qualifiers
1..1538
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/tissue_type="leaves"
/notes="ecotype: Columbia"
1..1538
/gene="RDRP"
<1..>1538
/gene="RDRP"
/codon_start=1
/evidence=experimental
/product="RNA-directed RNA polymerase"

```

ORIGIN

```

Alignment Scores: 3,34e-13 Length: 1538
Pred. No.: 266.50 Matches: 121
Score: 266.50 Conservative: 83
Percent Similarity: 38.06% Mismatches: 210
Best Local Similarity: 22.57% Indels: 123
Query Match: 8.85% Gaps: 17
Db: 8

US-09-913-878a-2_COPY_710_1282 (1-573) x ATH011977 (1-1538)
Qy 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db 175 GGAATAGGAAGATATCAGCTGAATTTGCTAGAAAGTAGCAGAGGAATGTGCTTAC 234
Qy 21 AspVal---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMet-TrpValI 39
Db 235 GAGTTTTCATCTGCTTTTCAATCCGTTATGGCGGTTATTAAGAGATGGTGGCTGT 294
Qy 39 eAspValAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTr 59
Db 295 G----- 295
Qy 59 pGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerG 79
Db 296 -----ATCCAAACTCATCAAGAAA-CTGCTCTCAGGAGAGATGAGCAA 341
Qy 79 uLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeu-----ProValle 93
Db 342 ATTCGAATCGGAGAACACCAAGCTTGTGCTGGCGTGGAGCAAGTACCACCTTGT 401
Qy 93 uGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLea 113
Db 402 TATGACACAGACAACCTGATTACACTTTTGTCTACTCTTGGAGTTACCGACAGTGT 461
Qy 113 nAspLeuGlnArgGlnPheSerGluGlnLysHisAla---LeuAsnArgProValGlu 132
Db 462 GAAGAAACAAGGAGAGTTGTGATGCTCTCGACGCAATCTTACTCATCTTTGGAA-- 519
Qy 132 eArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 152
Db 520 -----GTCACAGAGGCTCT 533
Qy 152 lProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSe 172
Db 534 TGGTTTAATGCT-----CCAGGGGAAACACAAATATTCTCAAGCATTTGATCTGT 587
Qy 172 rGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysAr 192
Db 588 TGGCTATAAAACCTGACGCTGAACTTCTTCTTCAATGATGCTTCAGAAATTCAGGGCATC 647
Qy 192 gLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTy 212
Db 648 TAAAGTTGTTGAACTACGGAACAAAACAGGATTTTCTTCGTGGGAAGATCTATGAT 707
Qy 212 rMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSe 232
Db 708 GGGATGCTAGACAGACAGACAGGCTGGAGTAGTATGGTCAGGTAGTAGTCAGTATTGAGA 767

```

```

/protein_id="CAA09894.1"
/db_xref="GI:4375829"
/db_xref="GO:092RX5"
/db_xref="SPTREMBL:Q92RX5"
/translation="IRAWMGDFDHNRVAKYARLQGSFSSSRRETLNVRSDIEVLPD
VRIISLGRYVFSFGIKISAFKARVAKCGLTFSPSAFQIRYGVYGVVAVDENS
SKKLSRKSMSKPESENTKDLVAMSKYQFCYNNRQLITLLSTLGTVDVSVFKKREV
VDRDLAILTHLEAEALGLMAPGENTNLKALILCGYKPDAPFLSMMLQNPFRASKL
LELRKTRIFISGSSSMSCLEDETLEYGQVVQVSDPMRPPRFITGPVVVAKNP
CHSGDVRVLQAVNPALNMVDCUVFPQKGLRHPNECSGSDLDGDIYFVCDQDLV
PRTSPEDYTTPETIOLDHVTIEIEEFANYINVDLSGLIARIAHTAFADKEPLKA
FDPCPCILAKKFSTAVDFPKTGVAAVIPQRLVKVEYDFMEKPDKFTYKSNVIGKLF
REVKERAPPLIISIKSTFLDVASKSYDKDMEVDGFEYVDEAFYQKANYDFKLGNLMDY
YGIKT"

```

| | | | |
|------------|--|--|------------|
| QY | 232 | rLysPheArgAspGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAl | 252 |
| Db | 768 | TCCCATGAGCGCGGGAAGCGA-----TTCATCATCACCAGCGCTGTGTGTGC | 818 |
| QY | 252 | aArgSerProAlaHisPheProSerAspIleGlnArgValAlaValPheLysProG1 | 272 |
| Db | 819 | CAAAACCCATGCTGCATCTCGTACGTCGGTCTTCAAGCTGCAATGTCACG | 878 |
| QY | 272 | uLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAlaL | 292 |
| Db | 879 | TTTAAATCATGCTGCATCTCGTCTCCGCAAAAGGCTTGAGGCCACACCCAA | 938 |
| QY | 292 | sLysLeuSerGlyGlyAspValAspGlyAspMetAlaTrpValCysTrpAspProGlu1 | 312 |
| Db | 939 | TGAATGCTTCGGAGTGATTAGATGAGATATTACTTTGATGTTGGATCAAGAAT | 998 |
| QY | 312 | eValAspGlyPheValAlaAlaGluMetProLeuGluProAspLeuSerArgTyrLeuL | 332 |
| Db | 999 | GGTTCGCCCAAGCTGCAACCAATGACTACCTCTGAACCACTCAATCTTGA | 1058 |
| QY | 332 | sLysAspLysThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaL | 352 |
| Db | 1059 | TCATGAT----- | 1065 |
| QY | 352 | sGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPh | 372 |
| Db | 1066 | -----GTCACAAATGAGGAATTCGAGAGTACTTTGCGAAGTACATTGTGAATAG | 1118 |
| QY | 372 | eLeuGlyMetCysThrAsnTyr-----LysGluArgLeuCysTyrI1 | 386 |
| Db | 1119 | TTTAGGATCATCGCAATGCTCATCCGCTTTCGGGATAAGGACCACTC----- | 1170 |
| QY | 386 | eAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuVa | 406 |
| Db | 1171 | -----AAGCGCTTAGTGACCCATGCTGAGCTTGCAAGAGTTTCACTGCCGT | 1223 |
| QY | 406 | lAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg1 | 426 |
| Db | 1224 | AGACTTTCCAAAACCTGGTGTGCA-----GCTGTGATACCGCAACA | 1265 |
| QY | 426 | uLeu-----LeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysase | 441 |
| Db | 1266 | TCATTAGCTGAAGAGTATCTGATTCATGGAGAACCGGATAGCCACATATGATC | 1325 |
| QY | 441 | rAspSerTrpLeuGly-----ArgGlyGluProThrHisI1 | 453 |
| Db | 1326 | GAAGAAGCTGATGTTGTAAGCTTTTAGAGAGGTAAAGAGCGAGCTCCACCGTGTATCTC | 1385 |
| QY | 453 | eLysAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu----- | 470 |
| Db | 1386 | GATCAATCATTCACGCTTGATGCGCTCAAGCTCTTATGATAAGACATGAAGTTGA | 1445 |
| QY | 471 | -----GluAlaPheHisAsnAlaMetLysAlaAlaLysAspTh | 483 |
| Db | 1446 | CGGATTTCGAGAGTATGTTGATGAAGCTTTCTAC-----CAGAAGCGCAATTATGATT | 1499 |
| QY | 483 | rGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyr | 498 |
| Db | 1500 | CAAGTTAGGT-----AATTGTGAGCTACTAT | 1527 |
| RESULT 7 | | | |
| LOCUS | AF468822 | 6688 bp | DNA linear |
| DEFINITION | Diaporthe perijuncta putative RNA-dependent RNA polymerase RDP-1 | | |
| VERSION | AF468822 | | |
| KEYWORDS | AF468822.1 | GI:18656896 | |
| SOURCE | Diaporthe perijuncta | | |
| ORGANISM | Diaporthe perijuncta | | |
| | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; | | |
| | Sordariomycetidae; Diaporthales; Valsaceae; Diaporthe. | | |

| | |
|-----------|---|
| REFERENCE | 1 (bases 1 to 6688) |
| AUTHORS | Preisig,O. and Wingfield,M.J. |
| TITLE | Putative fungal RNA-dependent RNA polymerase |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 6688) |
| AUTHORS | Preisig,O. and Wingfield,M.J. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (15-JAN-2002) Forestry and Agricultural Biotechnology Institute, University of Pretoria, Lumon Street 74, Pretoria 0002, South Africa |
| FEATURES | Location/Qualifiers |
| source | 1. 6688 |
| | /organism="Diaporthe perijuncta" |
| | /mol_type="genomic DNA" |
| | /isolate="CMW3407" |
| | /db_xref="taxon:186170" |
| | /country="South Africa" |
| | join(<1. 275,348. .>381) |
| | /product="unknown" |
| | join(<1. 275,348. .381) |
| | /notes="similar to D-lactate dehydrogenase (D-lactate ferri cytochrome C oxidoreductase)" |
| | /codon_start=1 |
| | /product="unknown" |
| | /protein_id="AAL78033.1" |
| | /db_xref="GI:18656897" |
| | /translation="MDDLGLFASILGHIGDNFHEILYNPNVNGEREKVEECVKQWVR |
| | RAIDMDGTCTGHEHSGIKGKDSLLLEVGPETLAVMRAIKGSLDPRWIMNPKIFDRQS |
| | " |
| gene | <1361. .>5870 |
| mrna | /gene="rdp-1" |
| | join(<1361. .4598,4654. .>5870) |
| | /gene="rdp-1" |
| | /product="putative RNA-dependent RNA polymerase RDP-1" |
| | join(<1361. .4598,4654. .5870) |
| | /gene="rdp-1" |
| | /codon_start=1 |
| | /product="putative RNA-dependent RNA polymerase RDP-1" |
| | /protein_id="AAL78034.1" |
| | /db_xref="GI:18656899" |
| | /translation="MSHSTPTTPDRGSAQPSAQIHHTTSSIPDATAKDNPR |
| | ISTLRDAIADTSDASNRNTQPSVVRRLPSPMSHTNSGNVLGOKVLSGPIQGRS |
| | IQNASARTFGPRGHPNPNRPNRQPSQSLQQRSTAPLPAVLASAAQSLTI |
| | PDNLPTDQIRIWEISMTNHLNVNRVEIYENFDGRTGKRVIIPEPPIRWPTWSG |
| | PDNLITDRADSSLOSULETRPFPSTPCNRPAPQCFCKKLLPLKLVFGWASE |
| | SAMLMPPQSYEDVLLTKFSKQFLOFTVNLGRYCKAFKSIDFSHPTKLIWGRR |
| | ENGAALAMLLPSPKLYGRLPVSHDDEQTLWQENDLNWAKQTNISSEMKLRDLPAK |
| | EKOTWKSIEFRWTTYYLDMESLSEWESVRYLADYNVKILQVDFDLNPGLSNAW |
| | KLDNNGAEMTLLELGGTSQYHLDFFVRYQLVEICSHGLLSEYGVNAQFLKLSFETD |
| | RAWMLEGAFAEKQPFPEMPEDDPRILHYMPNARIPPYATLVRAVITPTTIYFKT |
| | PCVELTNRIILKYSDLNDRFLVQFTDEVTFGKIFSSQSNKDDNLTYSRVKPAARMQV |
| | IIGDRHYKFLAFNSQPRENGAFECETDHTVTCASIRNMMDPHRISVFKPAPRMGO |
| | CFITTRVNGISIPKIQGDIORSGDNIWNTDVGKISVFARMIASERDLPTP |
| | SCFMRMGCGKGVVVVNPDPANEVHIFRSQEKFAVYNGLEILIKTSTFNATLKVQ |
| | IPLVIALGVDDAVFVRLMDLKEDEALDSMKAGELLRSQVDENQTLITMAEMVDT |
| | FMDSEEPFLTLRLKWKVLRKQKFAISVKNSAMIFGVVDETPTGRHGSQDTSGK |
| | GFNNIESLPQIFLQVPIEGSDGKSTTNYEITGICVVRNPNALHPGDVRYEADVPE |
| | LRHLNVVVPKTDQDRDPSMCGMDGDDYFYVWDERLIPSEMDHPDLHDADST |
| | PGLDSPADVTNGDVTFFAQMKNDSIGRIATAFAAQADOLAGVHKPKIEHLAKHS |
| | XADYIKSGPAVMKRLQPRNPMHMERDKSSVHSRGALGQIYDKIKTEBPHAAVE |
| | MSFDRILRSYQLEEDTLAKASIKKAYDIAMRLTGQHEAPTEFIMSTFILTKEP |
| | VGSDYKQENVGEMAAKARFVAMEAVTQDEPTPCASAKVNAVALGKLDPRFAAM |
| | YTTVTHNVRAAVRERKPRDDDELSIQMPLISPLWYHRELNRVALGKGVDRPL |
| | RKPCVADPAANIESESPSGDALSIEDQGETQLSLQVRSKDGKGVYIDRGSEVETG |
| | KVQEDYVYRTSSGOVVRGQLTLFTDTGEEVQGAQSTIATPFPANNNSGNPLVEDDG |
| | TGSSYPASPRSEHAEGEDEIEIFBEVDGANDGEDEALEILARKIQM" |
| | <6068. .>6526 |
| | /product="unknown" |
| | 6068. .6526 |
| | /codon_start=1 |
| | /product="unknown" |
| | /protein_id="AAL78035.1" |
| | /db_xref="GI:18656899" |
| mrna | |
| cds | |

/translation="MVLONHOQANRPTEFLSLTYSRASDYLISLVNAGDCPFLTWGHH
MMQCPAPQLGTSPTSPDRDPPFHQQRDSHRAPFQAPGTSIDLVDWIGLESABEP
AKLISLSLSIHLSPSRPSTMDRAVSSSLTADDSOKVAHQCFMLDPF"

ORIGIN

Alignment Scores:
Pred. No.: 2,84e-12 Length: 6688
Score: 265.50 Matches: 137
Percent Similarity: 38.2% Conservative: 67
Best Local Similarity: 25.70% Mismatches: 199
Query Match: 8.82% Indels: 131
DB: 8 Gaps: 23

US-09-913-878a-2_COPY_710_1282 (1-573) x AF468822 (1-6688)

Qy 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db 3512 GAGTGGGCAAGATACCTCTCTTTCAGAAAGATTCGACGAGCGGTATTTGCCA 3571
Qy 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db 3572 GAGCGCCTTCATGCTTTCAGATGAGATGGCGGGTCAAGGGCGTGTCTGTATGG 3631
Qy 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgGlyTrpGlu 60
Db 3632 CTGACATCCCGAAGCAGAG-----GTCCATATACGACCTTCTCAGAGAAAGTTTAA 3685
Qy 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db 3686 GCGGTGTAC-----AATGGTCTCGAGATCATCAAG--ACCTTACGTTTC 3727
Qy 81 LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys 100
Db 3728 GCCCATGCTACTCTCAATAAGCAGGTTATACCTGTCTGATCGCCCTGGGTGTGGAT--- 3784
Qy 101 ValLysMetArgGluAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db 3785 -----GAGCGCTTTCTCAGAAATGCTCGATGATGAGCTCAAAGATCAGAC 3832
Qy 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db 3833 GAA-----GCCTTA-----GCTGAT 3847
Qy 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db 3848 TCAATGAAGCCGGTGAATCTTCGCTCTCAAGTC-----GAC 3886
Qy 161 SerGlnGluThrLeuAsnPheLeu---MetAsnSerGlyPheAspProLysLysGln 179
Db 3887 GAGAACCAACACATTGACCATGCGGAGATGGTCGACACGTTTCATGCTCCGACGAG 3946
Qy 180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArg---LysCys-----Asp 195
Db 3947 CCGTTTCCTC-----TGGACCTTGCTTCGTTTGTGGAAAGTGTGGGTGTGAAG 3994
Qy 196 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAla 215
Db 3995 AGGTTAAAGCAAAAGTTCGCCATCAGCGTCAAGACACGCTATGATCTTCGGGGTGTGTC 4054
Qy 216 AspPheTrpGlyValLeuGluAsn----- 224
Db 4055 GATGAGACCGGTGTGTCGAGGGGCCACTCAAGACACCGAAGGCAAGGGTTCATATAAC 4114
Qy 225 -----GluValHisVal---GlyPheSerSerIlePhe 234
Db 4115 ATCGAGTCGCTCCACAGATCTTTCCTCAAGTCCCTATCGAGGGCAGTGATGGGAAGAGC 4174
Qy 235 ArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSer 254
Db 4175 ACCAGCAATACGAGGTATCATCTGGAAT-----TGGGTAGTAGGCGGTAAAC 4222
Qy 255 ProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHis 274

Db 4223 CCAGCGCTTCATCTCGAGATGTCGGGTGTTGAGGCGAGTTCACGTCGCCAGATTACGA 4282
Qy 275 SerLeuLysAspValIlelePheSerThrLysGlyAspValProLeuLeuAlaLysLysLeu 294
Db 4283 CATCTCAAGAAACGTGCTCTTTCAAAAGACTGGAGACAGGACGTTCCGACATGTGC 4342
Qy 295 SerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal--- 313
Db 4343 TCCGTTGGTGACATGACGAGGATGACTTCTTTGTTTACTGGGACGAGGACTCATCCCA 4402
Qy 314 -----AspGlyPheValAsnAlaGluMetPro----- 322
Db 4403 AGCGAGTGGGACCCACCCACTCTCGGACACGATCGGATTCATCGACTCTCGGCTGGAC 4462
Qy 323 -----LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThr 337
Db 4463 AGCCCGCGAGATGTGACAAACGGGAGCGTCACTAGGTTC----- 4501
Qy 338 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGlnInThrThrTyr 357
Db 4502 TTTGCCCGAGTATTAAGAAGCAGTATCCCTTGGCCGTATCGCC----- 4543
Qy 358 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGly----- 374
Db 4544 -----ACAGCGCATTGCTCAGGCGAGATCAGTTGCGAGGGGGCGTCAAG 4588
Qy 375 -----MetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsn 392
Db 4589 CACCCAAATGTAAATAATCGTACGACTTGTGTGGC--CGGGGTGAAGTACGGCT 4646
Qy 393 LysProAlaIleleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly 412
Db 4647 GAATAGGATCGAGTCGCCAAGCTCCACTCAAAAGCTGTGCTATATATCAAGTCCGCG 4706
Qy 413 IleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuGlyGlyAlaLeu 432
Db 4707 -----AAACCCGGCTCATGAGAGACATCTT-----CAGCTCGA 4742
Qy 433 SerLeuProAspProMetTyr-----LysSerAspSerTrpLeuGlyArgGlyGluPro 450
Db 4743 AACTGCGCGCATTTGAGTGGAGGGGCAAGAAAGTCTCTCCTACCTCCAGAGCGCGCTG 4802
Qy 451 ThrHisIleleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
Db 4803 GGTCAATTTACACCGGATCAAG-----ATC 4829
Qy 471 GluAlaPheHisAsnAlaMetLysAlaAlaLysAspThr 483
Db 4830 GAGCAATTCATCGCGCTTACGAAATGTCTTTCGACACG 4868

RESULT 8

DDI314910 5992 bp DNA linear INV 14-JUN-2001
LOCUS Dictyostelium discoideum partial rrpB gene for putative RNA
DEFINITION dependent RNA polymerase, exons 1-2.
ACCESSION AJ314910 GI:14475570
VERSION RNA dependent RNA polymerase; rrpB gene.
KEYWORDS Dictyostelium discoideum
SOURCE Dictyostelium discoideum
ORGANISM Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1
AUTHORS Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P.
and Nellen, W.
TITLE RNAi in Dictyostelium: developmental regulation and the role of
RDRPs and dRNAse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5992)
AUTHORS Martens, H.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2001) Martens H., Genetics Department, Kassel
University, Heinrich-Plett Strasse 40, 34132, GERMANY
FEATURES Location/Qualifiers

```

source
1. .5992
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
1. .5992
/gene="RrpB"
/join("<1.970,1067. .5992)
/join("<1.970,1067. .5992)
/join("<1.970,1067. .5992)
/join("<1.970,1067. .5992)
/codon_start=2
/product="putative RNA dependent RNA polymerase"
/protein_id="CAC41975.1"
/db_xref="GI:14475571"
/db_xref="GOA:Q95266"
/db_xref="SPTRMBL:Q95266"
/translation="RTIETVQRSSKDEMRMLSRVFKPHSLTNTSTSEIQLPELVSF
KTSQESLIETICEFLVNSKELQIFLNSVDLDGVRGNSFLKALERDKLSKQL
ESIYETVQLVILKELLSVLTGEPKQVLTGLDLMOSTNOENELYNKLLELYVIS
GOYEDRESNFKSTYRKLISLVAISDASDGLQDLRLVLFVETRSGASNTSM
LKEPEOEVLYHVKRLVGHNGDDGMOSEKQSIIRKFRQCKELIVTVTBEGLDVOD
CNVICDGLSLSKLQRRGRARSKNESKPIIYNDKDKLILNLSLSENLLNSIN
QILSNRERQATNQFNKLWKNINNEYFINNQLDGGPLDKIDKCVLSIYHIEDIE
TFRELLFQPLDVNQVQVSIIRMEGDDYLSSTIIVILNTCIDKEQSIYRLKSLNI
NGLYWMIIDRSFHEFDNDNDNNNSYDRENOFNFFINEYTHLSQFKNKYSI
GSPYGNMSPSTKYHIGIQLDSDSFDLFGNREMFYKTRFORVRLNKLTDIESTFL
LVNDKYONTVHILLIYKREPLIPEEKETTHRVFNKSKPKDDFQCAFVYRLTSLTD
NETKEVILKPLANGPEMYESSILELNCFIQGGINKNNCMWIDLNQYIKRNLETHYLL
QCLQSSQSVGRVLNQDFIVEINNLIAFGQBEAETILSLSKRTKFFINYEKFGDCL
KNLPDFTYAKKDLSTKSGEFTMINDVFETSRFTIFQPTLQRCRAVRKFGSNFI
MKVIVNESPELQGVNKHPLHENRIKPILEGGLVGGKIYSYAGNSQREVSWF
VNSQIETHVTKWNGIEHVDNVRKFCRIGLMFSTTPTVLPQNKRIYIQIDTRNTH
VFTEGGEIGELAKLHENYFRSTCAQVRIQGNKGLVNNQAPDPSGIYRPS
MKVFNIVCGDEHRTLEICSVSTSRCKLNROVIGLLSTGLQDNNVFALODHYLNQV
AOIVNDTASKQAIPEFFEDITEGELYQDPYIRRLISLYLKLKMERIOQKCHIEKDS
RMLLVGCDPTNSLPNPTVEQLEEDDDGCKYKVEGLEGLVWIKNCPHGDVRY
LKAQNLRLHRLHNLVSTKGDVENFEISGDDGDRYFCYDKSLIGRASKETA
YIGDETNSNDKKAIVNDFPFAALSNYSTNVERQELGKYNHSHLALSDFGANHEFSI
QSKCEKFEIDYPTKGTIGHTIPKANVWLTKGYPHYQWRENSRVYQSKTIMGMY
DQDQILVYIGDFLPISLDKSNLVDDEIYLSNAKILYSQKLVHSLRHSASEE
SIMGFLDQGFSDKYSKDIKGEMKNDYIKIQTPENBFLKEFGQHKENCILLIHRVN
IEKKVSAWYHVAUSDLDKRALGFWIARSPKQMTLDOESOKKNILVNSILDHWNKX
IKLPLFYSRLNIVKVELJLPLANLNDTVFKESTLEIIGSTATMLFDNSNLNLYKL
ESQNDHMLAEKNEILNQLKDELKTSPLSBAECSILKIQGDILPFPKIKIQCILIG
NPMLNTTNSLTIGNWKLIEIASRFSVDNNNNNNKHKRKNKRLDEYNTNLLG
QKVLVENSMPISLNNKNNNNNNNNNNKHSVKNLTKSNLFSLLIDFCILNKYLDKC
DTLSLFRYSDFGRSIYGNBFFIILKFSKDANDIENQILNPSYESHTKALLFYQES
FRHAPSLSTFVSVERFLSNCLTKTKKIMPKINQINLLOERSVNEEIVKXNVLQGCQ
IKIDTVKNPYLIDGKPOQIRLAQNAIRDLLSIKTENYNLAIKRNKRNKFTIGDF
AKNKIK"
<1. .970
/gene="RrpB"
/number=1
971. .1066
/gene="RrpB"
/number=1
1067. .5992
/gene="RrpB"
/number=2
exon
intron
exon
ORIGIN
Alignment Scores:
Pred. No.: 4e-12 Length: 5992
Score: 263.00 Matches: 134
Percent Similarity: 36.53% Conservative: 91
Best Local Similarity: 21.75% Mismatches: 209
Query Match: 8.74% Indels: 182
DB: 3 Gaps: 25
us-09-913-878a-2_copy_710_1282 (1-573) x DDI314910 (1-5992)

```

```

QY 1 GlyValGlyArgMetSerArgSerValAlaValArgIleArgAspValLeuGlyLeuGly 20
DB 2852 GGATGTGTTGTAATTTGCTCCAGAACTTGTAAACATTTAAATGAAATTAATTTTCA 2911
QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaValGlyMetTrpValIleAsp 40
DB 2912 CCATCAACATGTCATATCAATAGATTAGATTAGATTAGATTAGATTAGATTAGATT 2971
QY 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
DB 2972 ---AATCAAGACCCGATCCATCTGTTTATATATTAGACCAAGTATGTTGTAATTC 3028
QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
DB 3029 CCAATCGTTTGGTGTAGTGAACATAGACATTAGACATTGTCAGTGT---TCAACA 3085
QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAsp---ArgAlaArgAsp 99
DB 3086 TCAAGATGTAATTAATTAAGCAAGTAATTAATTAATTAATTAATTAATTAATTA 3145
QY 100 LysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPhe 119
DB 3146 AATGTT-----TTCTTT 3157
QY 120 SerGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTrpGluSer 139
DB 3158 GCATTACAAGATCATTATCTAAAT-----CAATGGCTCAAAATCGTAAATGATACC 3208
QY 140 TyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuPro 159
DB 3209 AATGCTTCCAAACAGCAATCGTTGAATC-----TTTCCA 3244
QY 160 AspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGln 179
DB 3245 GATATTACAGAGGTGAACTT-----TATCAAGAT 3274
QY 180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSer 199
DB 3275 CCTTACATCTAGATAATTTAATGATCTTTATATAATTAATAATGAGAGAAATCAACA 3334
QY 200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
DB 3335 AATGTCTATAGAGATAGAAATAGATAGATGTTGTTGGGTGTTGTTGTTGTTGTTGTT 3394
QY 220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGlu 239
DB 3395 TCATTACCAACCAATCTGTTATTTGTTCAATTGGAAGAGAGATGAAGATGATGATGAT 3454
QY 240 -----SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
DB 3455 GGTCGTAATATGAAAAAGTAATTGAAGGATTCGTAATGTAATTAATAATCCTTGTACT 3514
QY 258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
DB 3515 CATCTCGTGTATGATGATATTTAAAGGCTGTTGATAATTTAAAGATTAAGACATTTAAGA 3574
QY 278 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 297
DB 3575 AATGTATTAGTTTTCACAAAGAGGTGACGTACCAAACTTTAAAGAAATTTTCAAGTTCA 3634
QY 298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluLeuValAspGlyPheVal 317
DB 3635 GATTTAGATGTTGATGCTGTTTCTTTTGTAGCATAAATCATGTGAT----- 3682
QY 318 AsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThr 337
DB 3682 ----- 3682
QY 338 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaLysGlnGlnThrThrTyr 357
DB 3683 -----GGTAATCGTTCAAAATCAGACAGACGCTAT 3712
QY 358 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGly----- 374

```

```
Db 3713 -----TTAGGTGACGAGACC 3727
Qy 375 MetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysPro 394
Db 3728 GTTCCAAACAAATGATAAAGGCC-----AATGTTTCAATGATCCA 3769
Qy 395 AlaIleLeuSerSerLeuValAspGlnSerLysGlnGlyIleVal 414
Db 3770 ---TTTGCTTTATCTCACTGACTCTACAAATGTTGAACCTCAAGAACTTGGTAAATG 3826
Qy 415 PheAsnGluAlaSerTrpAlaGlnLeuArgGluLeuLeuGlyGlyAla----- 431
Db 3827 TATAAT-----TCACATTTGGCAATCTCTGATCTTTTCGGTGCATAACAGAA 3874
Qy 432 -----LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGlu 449
Db 3875 TTTTCAATCAANTCCAAAGATGTTTAAAGAA----- 3910
Qy 450 ProThrHisIleAlaAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGlu 469
Db 3911 -----ATTGATTATCCAAAAAAGTGTATTT---CATGGCACTATTCCAAAAAGAA 3955
Qy 470 LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPhe 489
Db 3956 GCAAAATGTTTG-----TTAAAACTAAAGGTTATCCACATAT 3994
Qy 490 TrpAspProAspLeuAlaSer-----TyrTyr----- 498
Db 3995 ATGCAAGAGAGAAATCAACCAGAGTTTATTATCAATCGAAAAACCATTTATGGTAAATG 4054
Qy 499 -----ThrPhePheLysGluIleSer---Asp 506
Db 4055 TATGATCAAAATGATGACTGTTGTTATATGTTGATTTCTTACCATAATATTAGTCTTTCAT 4114
Qy 507 LysSer-----ArgSerSerAlaLeuLeuPheThr 516
Db 4115 AAAGCAATTTGGTGTGACTATCAATCATTTTAAATAGTGTCTAAATCTTATATAGT 4174
Qy 517 ThrLeuLeuAsnArgIleGlyGluValGluLysGluTyrGlyArg----- 531
Db 4175 CAGTATAAACTTCAAGTTTCATTTCATTGTTTAAAGACATTATTTCAGCGGAGTCTGAAGAATCT 4234
Qy 532 -----LeuValLysAsnLysGluMetArgAsp----- 540
Db 4235 AATATGATGTTTCTTGATCAAGGTTTATTAGTGACAAAGTTTCAAAAGATATCAAA 4294
Qy 541 -----SerLysAspProTyrProValArgValAsnGlnValTyrGlu 554
Db 4295 GGTGAATGAAAAATGATTAC---ATTAAATTCACAAACTTTTGAA 4339

RESULT 9
LOCUS AR145904
DEFINITION Sequence 1 from patent US 6218142.
ACCESSION AR145904
VERSION AR145904.1 GI:15109093
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
UNCLASSIFIED.
REFERENCE 1 (bases 1 to 3731)
AUTHORS Wassengger,M., Riedel,L., Schiebel,W. and Sanger,H.L.
TITLE Nucleic acid molecules encoding polypeptides having the enzymatic
activity of an RNA-directed RNA polymerase (RDRP)
JOURNAL Patent: US 6218142-A 1 17-APR-2001;
FEATURES Location/Qualifiers
source 1..3731
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
```

```
Pred. No.: 2 6e-12 Length: 3731
Score: 282.00 Matches: 127
Percent Similarity: 37.66% Conservatives: 85
Best Local Similarity: 22.56% Mismatches: 222
Query Match: 8.70% Indels: 130
DB: 6 Gaps: 17

US-09-913-878a-2_COPY_710_1282 (1-573) x AR145904 (1-3731)
Qy 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db 1772 GGAATTTGGTAAATATATCTGCTGACTTTGCTCATAGAGTTGCCTCAAAATGTGCGCTCAA 1831
Qy 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db 1832 TATACCCCATCTCTTCCAGATTCCTTATGTTGGATATAAAGGTGTTGCGGTGTGAT 1891
Qy 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db 1892 CCGATTTCATCAATGAAG-----TTGTCTTTTGAGAAAGAGCATGTCGAAATATGAA 1942
Qy 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db 1943 TCAGACCAACATA-----AAGTTAGATGTCCTTGA---TGGAGCAATAT 1984
Qy 81 LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys 100
Db 1985 CAGCCTTGTTATCTTAATCGTCACTGATTACGCTCTGCTCT----- 2026
Qy 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db 2027 -----ACACTGGAGTGAAGATGAAGTTCTCGAACACAGCAAGCAAGAGCTGTA 2077
Qy 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db 2078 GATCAGCTTGATGCTATCTTGCATGATCTTTGAAGGCACAGAGGCTTTGGAATGTATG 2137
Qy 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db 2138 TCT-----CCTGGA 2146
Qy 161 SerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
Db 2147 GAGACACATAATTTCTCAAGCAATGCTAAACTGTGGTTATAAGCCTGATCTGAGCC 2206
Qy 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db 2207 TTTCTTTCAATGATGTTCCAAACCTTCCGCGCATCCAAAGTTCTCGATTGCGGACTAGA 2266
Qy 201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
Db 2267 TCAAGATATTTATTTCCAAATCGAACAATGATGGGATGTTTGAATCAATCCAGAAC 2326
Qy 221 LeuGluGluAsnGluValHisValGlyPheSer-----SerLysPheArgAsp 236
Db 2327 TTGGAATATGTCAGGTGTTGTTTCAGTTTACTGTTGCTGGACATGAGAGATTTTCTGAC 2386
Qy 237 GluGluGluSerPhe-----ThrLeuLeuSerAspCysAsp 248
Db 2387 GATTTATACCTCAATTTAATACAGCAGATCCACCAAGATATTTCTTCTGAGGGAAT 2446
Qy 249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
Db 2447 GTGGTTGTTGCAAAAAATCCATGCTTGATCTCGTGATATTCGTGTTTAAAGGCTGTA 2506
Qy 269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
Db 2507 AATGTTTCGAGCGCTGCACCACATGTAGATTGTGTTTATTCCTCAGAAAGAAAGAA 2566
Qy 289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
Db 2567 CTTATCCGAATGAATGTTCTGGAGTATTGAGTGGGATATCTACTTTCTTTGCTGG 2626
```



```

QY 309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
Db 2627 GATCAAGACATGATCCGCCAAGCAAGTCCAGCGGATGGAATATCTCCACCAACCCAGC 2686
QY 329 ArgTyrLeuLysAspGlyThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
Db 2687 ATACAGTTGGACCATGAT----- 2704
QY 349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
Db 2705 -----GTCAATTTGAGGAAGTTGAAGACTCTTCCACCACTATATT 2746
QY 369 GlnProAsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuLys 384
Db 2747 GTCAATGACAGTTGGGAATCATCAAAATGCCATGCTGATTTGGCAGACA----- 2800
QY 385 TyrIleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsn 404
Db 2801 ---GAACTGATGATGCCATGATGATCCATGCAAAACCTTCTGACCTCTTTCAATT 2857
QY 405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db 2858 GCAGTGGAGCTTTCCAAAGACTGCTGCTCCCGCTGAAATACCATCT-----CAGTTGGC 2911
QY 425 ArgGluLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
Db 2912 -----CTAAAGAAATACCCAGACTTCATG 2935
QY 444 PheGlyArgGlyGluProThrHisIleLeuAspTyrLeuLysPheSerIleAlaArgPr 464
Db 2936 GATAAGCCG-----ACAAGACC 2953
QY 464 AlaIleAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
Db 2954 AGCTATATCTCAAGAGACTTATTGGAAGCTTTTC-AGGAAGTGAAG----- 3001
QY 482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLys 502
Db 3002 -----GACAAAGCA-----CCTCAGGCTAGCTCTATCGGACCTTCACAAAG 3042
QY 502 sGluIleAspLysSerArgSerSerAlaLeu----- 513
Db 3043 AGATGTTGCAAGAGATCATATGATGCTGATGATGAAGTTGATGATTGAAGATTACAT 3102
QY 514 -----LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysG 528
Db 3103 TCACGAAGCTTTTGACTACAAACTGAATATGACAAACAGCTGGTGAATTTAATGGACTA 3162
QY 528 uTyrGly 530
Db 3163 CTATGGC 3169

RESULT 10
LEDRP
LOCUS 3731 bp mRNA linear PLN 18-DEC-1998
DEFINITION L.esculentum mRNA for RNA-directed RNA polymerase.
ACCESSION Y10403
VERSION Y10403.1 GI:4038591
KEYWORDS RDRP gene; RNA-directed RNA polymerase.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1
REFERENCE
AUTHORS Schiebel,W., Pelissier,T., Riedel,L., Thalmair,S., Schiebel,R.,
Kemp,D., Lottspeich,F., Sanger,H.L. and Wassenegger,M.
TITLE Isolation of an RNA-directed RNA polymerase-specific cDNA clone
from tomato
JOURNAL Plant Cell 10 (12), 2087-2101 (1998)
MEDLINE 99055198
PUBMED 9836747
REFERENCE 2 (bases 1 to 3731)

```

```

AUTHORS Wassenegger,M.
DIRECT SUBMISSION
TITLE Submitted (08-JAN-1997) M. Wassenegger, Max-Planck-inst. fuer
JOURNAL Biochemie, Viridiforschung, Am Klopferspitz 18a,
Planegg-Martinsried, 82152, FRG
FEATURES
source
1..3731
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/strain="Rutgers"
/db_xref="taxon:4081"
1..3731
/gene="RDRP"
/EC_number="2.7.7.48"
/codon_start=1
/product="RNA-directed RNA polymerase"
/protein_id="CAA71421.1"
/db_xref="GI:4038592"
/db_xref="GOA:Q9ZR58"
/db_xref="SPTREMBL:Q9ZR58"
/translation="MGKTIQVGFPPYLSSAEVVKSFLEKTYGTGTVCALELVQSKGGS
RAFAKQFADNISADKIITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDMQDITLNF
GQISDDKFAVLGSEVSIQFGLKLPFFLLSSGADYKLLQSYENIMQVVLHRRPYG
QNAQFLLIOLFGAPRIYKLENSCYSPFKETPDQWVTTDPFPPWIGISSLICLPFR
RGVRLNPFESPEHYAERENNTLOTGFTTFVYSKALVENVQPEGIGISPYKILFKI
SLVQHGCTPGPALAVYPERLVDPERRVACIEHALEKLYIKCCYDPMVRLTEQYD
GILKQKPPSPSITLDDGLYVRRVLTVPCKYFCGPEVNSVNLNYSNEDIONFL
RVSFDEWEKLYSTDLLPSTGSGVRNTRYERILSTLTKGFIQDGGKFFELARSSS
QURDNSVMFASRPGLTANDIRAMGDFSQIKNAVYAARLQGSFGSSFETLSVLNKE
IEVTPDKVHGTSYVFSIGIKISGFDAHRVASKGLQVTPSAFQIRYGGYGVGVGD
PDSMKLSRKMSKYSYSDNIKLDVLGWSKYQCYLNRLQITLTLSTLGVKDEVLQEQ
KEADVDLAILHDSLKAQALEMSPGENTNLIKAMNGCYKPDAPFPF.SMWLQTPRA
SKLLDLETSRIFIPNGFTMGCLDESRLYGVQVFOFTGAGHGEFSDLDLHPFNRS
STNSNFIKGNVVAKNPCLHPGDIRLVKAVNRALHENVDCVVPQKGRPHNRS
GSDLDGDIIFVCDQDMIPRQVPMPIYFPAFISQIDHDVTIEVEEYETNTIYNDSL
GIANAHVVFADREPDMAISDFCKLAELFSAVDFPKTGVEAEIPSQURKPEYPDFM
DKPDKTSYISERVIGLFRKVKDKAPQASSIATFTDRVAARSYDADMEVDGFEDYIDE
AFDYKTYDNKLGNDMDYYGITEABILSGIMKASKTFDRDKDAEIAISVAVRALRKE
ARAWPKRNDIDDMLPKASAWVHTVHTYWGYNQGLKRAHFI8PPMWCYVDQLGIK
KOKARNRPVNLSSLEAQLSHRLVLK"
ORIGIN
Alignment Scores: 2.6e-12 Length: 3731
Pred. No.: 127 Matches: 127
Score: 262.00 Conservative: 85
Percent Similarity: 37.66% Mismatches: 222
Best Local Similarity: 22.56% Indels: 130
Query Match: 8.70% Gaps: 17
DB: 1
US-09-913-878A-2_COPY_710_1282 (1-573) x LEDRP (1-3731)
QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db 1772 GGATGTTAAATAATCTGGTACTTGTCTCATAGATTGCCTCAAAATGTGCCCTTCAA 1831
QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db 1832 TATACCCCATCTGCTTCCAGATTCTGATGATGATGATGATGATGATGATGATGAT 1891
QY 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db 1892 CCGATTTCATCATGAAG-----TTGCTTTGAGAAGAGCATGTCGAAATATGAA 1942
QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db 1943 TCAGACCAACATA-----AAGTTAGATGTCCTTGGA---TGGAGCAATAT 1984
QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys 100
Db 1985 CAGCTTGTATTCTTAACTGTCACATGATTACGCTCTTGTCT----- 2026

```



```
QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db 2027 -----ACACTTGGAGTGAAGATGAAGTCTTCGAACAGACAAAGAGAGCTGTA 2077
QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrgLysSerTyr 140
Db 2078 GATCAGCTTGATGCTATCTTGATGCTTTGTAAGGACAGGAGGCTTTGGAAATTGATG 2137
QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db 2138 TCT-----CCTGGA 2146
QY 161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLys 180
Db 2147 GAGAACACTAATATTCTCAAGCAATGCTAAATCTGTTATTAACCTGATGCTGAGGCC 2206
QY 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db 2207 TTTCTTTCAATGTTGCAACCTTCGCGCATCCAGTTGCTCGATTTCGGGACTAGA 2266
QY 201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
Db 2267 TCAAGAATAATTTATTCGAATGGAAGAACAAATGATGGGATGTTTGGATGAATCCAGAAC 2326
QY 221 LeuGluGluAsnGluValHisValGlyPheSer-----SerLysPheArgAsp 236
Db 2327 TTGGATATGTCAGGTGTTGTTTCACTTACTGCTGGACATGGAGAGTTTCTGAC 2386
QY 237 GluGluGluSerPhe-----ThrLeuLeuSerAspCysAsp 248
Db 2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGATAATTTTCATCTTGAAGGAAAT 2446
QY 249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
Db 2447 GTGGTGTGTCAAAAAATCCATGCTTGCATCTCTGTTGATATTCGTTGTTTAAAGCTGTA 2506
QY 269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
Db 2507 AATGTTTCGAGCGTCGACACATGGTAGATTGTTGTTGTTATTCCTCCAGAAAGAAAAAGA 2566
QY 289 ProLeuAlaLysLysLeuSerGlyAspTrpAspGlyAspMetAlaTrpValCysTrp 308
Db 2567 CCTCATCCGAATGAATGTTCTGGAGTGAATTTGGATGGGGATATCTACTTTGTTGCTGG 2626
QY 309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
Db 2627 GATCAGACATGATCCCGCAAGGCAAGTCCAGCCCATGGAATATCCTCCAGACCCAGC 2686
QY 329 ArgTyLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
Db 2687 ATACAGTTGGACCATGAT----- 2704
QY 349 SerAlaAlaLysGluGlnThrThrTyArgMetIleGlnLysSerPheHisPheAlaLeu 368
Db 2705 -----GTCACAATTGAGGAGTTGAAGACTACTTACCACACTATATT 2746
QY 369 GlnProAsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCys 384
Db 2747 GTCAATGACAGATTGGGAATCATAGCAAAATGCCAATGCCCATGCTCGATTTCGACAGACA 2800
QY 385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db 2801 ---GAACCTGATATGCGCATGATCATCATCAAAAAAATCTGCTGAGCTCTTTCAATT 2857
QY 405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db 2858 GCAGTGGACTTTCCAAAGACTGCTGTTCCCGCTGAAATACCATCT-----CAGTTGGCC 2911
QY 425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyLysSerAsp-SerTr 444
Db 2912 -----CCTAAAGAAATACCCAGACTTCATG 2935
```

```
QY 444 pLeuGlyArgGlyGluProThrHisIleIleAspTyLeuLysPheSerIleAlaArgPr 464
Db 2936 GATAAGCCGG----- 2953
QY 464 cAlaIleAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
Db 2954 AGCTATATCTCAGAAAGAGCTTATTGGAAGCTTTTC-AGGAAAGTGAAG----- 3001
QY 482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyTyThrPhePheLys 502
Db 3002 -----GACAAAGCA-----CCTCAGGCTAGCTCTCTATCGGACCTTCAAG 3042
QY 502 sGluIleSerAspLysSerArgSerSerAlaLeu----- 513
Db 3043 AGATGTTGCAAGAGATCATATGATGCTGATGGAAGTTGATGGATTTGAAGATTACAT 3102
QY 514 -----LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysG1 528
Db 3103 TGACGAAGCTTTTCTACATAAACTGAATATGACAAACAGCTGGTAAATTTAATGGACTA 3162
QY 528 uTyGly 530
Db 3163 CTATGGC 3169
RESULT 11
DDI314909
LOCUS DDI314909
DEFINITION Dictyostelium discoideum partial RtpA gene for putative RNA dependent RNA polymerase, exons 1-3.
ACCESSION AJ314909
VERSION AJ314909.1 GI:14475568
KEYWORDS RNA dependent RNA polymerase; RtpA gene.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1
AUTHORS Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P. and Nellen,W.
TITLE RNAi in Dictyostelium: developmental regulation and the role of RRPps and dRNase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5731)
AUTHORS Martens,H.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2001) Martens H., Genetics Department, Kassel University, Heinrich-Plett Strasse 40, 34132, GERMANY
FEATURES
source
1..5731
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
1..5731
/gene="RtpA"
/join("1..1572,1655..>5731)
/gene="RtpA"
/join("1..1572,1655..>5731)
/gene="RtpA"
/functions="required for RNA interference"
/codon_start=1
/product="putative RNA dependent RNA polymerase"
/protein_id="CAC41974.1"
/db_xref="GI:14475569"
/db_xref="GOA:Q95ZG7"
/db_xref="SPRMBEL:Q95ZG7"
/translation="METKDVHYKTKLPQKPMIPKPEGVETIPRDYQIESVYQSTQ
DNTLLVFTGMKTLVIMTLEMFSDNDQNSGDSKRIALFLVDRPLVLTQAGAI
EATNLKVLKYEINDSRFAVRKEYDVLVSTVSGVNLVLEVRNLIDFYITF
DEVHATGEHDNFKVDYIRKTDINYPRIGLTASLVISGNSTIDIVQSIKDMER
MLSRVFKPTSLTNSTQSELOPVLSPKTSQBSLIETSCFPLVSNKELQIFLAYS
DVLDDVGRGNPSLKALERLDKYSKQESLYIEYTKVLIYALLSVLSTEGPKQVLT
RDLDMOSTQNEELYNKLELRYVLSGOVEDRESNEPKGSTVRKLIISLEVAISDA
SDGELQDLRLVFRVETFRGASNLITSMKKPEQYELTKRLVGHNGVDGMDSEKQSS
```



```

QY 470 LeuGluAlaPheHisAsnAlaMetIlyAlaAlaIyAspThrGluAspGlyAlaHisPhe 489
Db 4649 TATGAATCTATTAAATAGTCTAAATC-----ThrPhePheIySylurle 504
QY 490 TtpAspProAspLeuAlaSerTyTy-----TTATATAGTCAGTATAAAGTTCAGTTCATTGTTTAAACATTAT 4726
Db 4679 -----TTATATAGTCAGTATAAAGTTCAGTTCATTGTTTAAACATTAT 4726
QY 505 SerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGlu 524
Db 4727 TCAGCAGAATCTCAAGAAATCTATAATG-----ATTGGTTTC 4762
QY 525 ValGluLysGluTyGlyArgLeuValLysAsnLysGluMetArgAsp----- 540
Db 4763 CTGTGATCAAGTTTC-----ATTAGTGATAAAGTTCATAAGATATCAAGGTGAA 4813
QY 541 SerLysAspProTyProValArgValAsnGlnValTyGlu 554
Db 4814 ATCAAAATGATTAC---ATTAAATTCACAAACTTTTGAA 4852

RESULT 12
AP004357/c
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
BAC clone: B1074C08.
ACCESSION
AP004357 BAC00010
VERSION
AP004357.4 GI:21328130
KEYWORDS
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1
AUTHORS
Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
Katayose, Y., Wu, J., Nimura, Y., Cheng, Z., Nagamura, Y.,
Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y.,
Idonuma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakasima, M.,
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,
Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,
Zhong, H., Iwana, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.I., Eun, M.Y.,
Yano, M., Jiang, J. and Gojobori, T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)
2
PUBMED
12447438
TITLE
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (19-NOV-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jun 6, 2002 this sequence version replaced gi:20975426.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, ELASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(http://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBS accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level

```

such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of B1074C08 clone has an overlap with B1147B04 clone (DBJ:AP004641) at the position 1 to 108,335 of 5' end and an overlap with B1129H01 clone (DBJ: AP003370) at the position 115,348 to 165,701 of 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

```

source
1. 165701
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="B1074C08"
complement(join(1956..2373,2716..2783))
/genes="B1074C08.1"
complement(join(1956..2373,2716..2783))
/genes="B1074C08.1"
/notes="contains EST D22801(C1352)"
unknown protein
/codon_start=1
/protein_id="BAC00710.1"
/db_xref="GI:21328131"
/translation="MDSTTTSQKATATATFITSTVIRNRLGRTSGATKKPKGASFQ
AMESYLVTFIVFAFSDSMVIVQTHSLTLEKSYCYCYLHFFLLASNYKPT
PFWHSQPKKNQPPRRSDPTNEATNRPTEDRGMLEDVGVGDQSQRSSMTQRWM
M"
join(7311..7428,7640..7776)
/genes="B1074C08.2"
join(7311..7428,7640..7776)
/genes="B1074C08.2"
/notes="contains EST AU069673(C0828)"
unknown protein
/codon_start=1
/protein_id="BAC00711.1"
/db_xref="GI:21328132"
/translation="MHCHVCVNCVNSGYAQOQFFVACVRAAQPRKHVSNTRPAARM
KLIGSGSGGQRRGIARALKRKHRIYIRCVVMLRC"
join(9087..9169,9577..9836,9998..10158,10204..10298,
10342..10570)
/genes="B1074C08.3"
join(9087..9169,9577..9836,9998..10158,10204..10298,
10342..10570)
/genes="B1074C08.3"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAC00712.1"
/db_xref="GI:21328133"
/translation="NARQCLDRVGRDGAANAVEKREVEVTDMGPPVGFPHLP
PLPKPATTAPTPLAVPAPRLHSSTPAPSPRLPSLPAHAPKRAQFDLNP
PLISLSLPHVGPSPRAVAFGACRAISSTSLSKSAIGSHSCAPGPFHRR
RPFWSLTVLELPSLPFGKRSFLFAGVFFHAGRIAPRAEPRRSHRGLRPLPP
PAPPVTRAAASHGLRVFEAPSPAPRAVRSARGSGSGSCGPAVDHPRGPGP"
join(16765..17574,19784..20080)
/genes="B1074C08.4"
join(16765..17574,19784..20080)
/genes="B1074C08.4"
/codon_start=1
/products="putative DNA damage repair protein"
/protein_id="BAC00713.1"
/db_xref="GI:21328134"
/translation="MLGGLYGLDLPSPSSADDDKPSAAGSSAAKVAPTLRKPATF
APPVSVLRNSRPAPKAPAAQPPPTLPITETTSFQFALVAVQSTVMEYDPPRP
DYEDYKDKLKRKAEEVKELERREERERELREGERDALNLSGEEAKRR
AMSSGAAPRSSPPHGGFAIGNSSGLGIGAGGOMTAAQRMWAGKGGGLG
KROGHTALVAKKDRRGVIVDENSSQKPKSVNFDGPTVTLIRNMGVEV
DDELEEVASECSKYGTDLRVLIFFITQDFFAEAVRIFILFAEEATKAWIDEG

```

RFFGGRVATFFDDEBFQKQLAPMPGEVAGFD"
 join(22181..22363,22471..22640,22731..22803,22919..22993,
 23119..23229,23337..23382,23475..23538,24200..24302,
 24389..24484,24576..24605,24824..24923,25151..25235,
 25355..25478)
 /gene="B1074C08.5"
 /CDS
 join(22181..22363,22471..22640,22731..22803,22919..22993,
 23119..23229,23337..23382,23475..23538,24200..24302,
 24389..24484,24576..24605,24824..24923,25151..25235,
 25355..25478)
 /gene="B1074C08.5"
 /note="contains ESTs"
 C36310(C32073),DLS391(C0569),AUI00694(C12073)
 unknown protein
 /codon_start=1
 /protein_id="BAC00714.1"
 /db_xref="GI:21328135"
 /translation="WAPGTTAPKRRKABSPSPSPMGDSGCGYSDSLDHAEBES
 FYSARSGDDQVSSNDDDDSEEEQEREMDEDEEDDDDDDEEMNEDEDEGE
 MELEKEYTLTQNOILLETIKQHRDDDVSKQAVKQKVLWRLERLLOKAFS
 TGNLKEPIRMFCQHNQIEQAVLDLLNSKQTLGSMELQEAALLERNRATKDVTD
 TNSLNGEDDEWSEVQKLRITPFNRSEIDKWKQTVQVTTGAALJKLHAFNPD
 IDQVTSYRDPFRMINRMLRSTLGVGESEVGHENNEENTEGDPFLVDSEFY
 QOLLKEFLSCDAGASEAFYALKQKHKKLVDRRASKSRKIRYHVHEKIANFMAP
 VPMVPPMAPKLFENLFGMGNOKSTTA"
 join(26433..26525,30039..30091,31996..32168,32606..32724)
 /gene="B1074C08.6"
 join(26433..26525,30039..30091,31996..32168,32606..32724)
 /gene="B1074C08.6"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAC00715.1"
 /db_xref="GI:21328136"
 /translation="WEIHTKLLDAAOLAVLDRRVFVYTCMKPSSSSFWVDITSP
 LQHRQVTELLLAIVSRGLAHFFHRDPQGLSTPAKSAWMLTPDVQQAHTIAHLTQ
 LSHQVSTTEKILIGAWEPALGAFNRRHGGDTDDFOSRCRF"
 36236..36550
 /gene="B1074C08.7"
 36236..36550
 /gene="B1074C08.7"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAC00716.1"
 /db_xref="GI:21328137"
 /translation="WAAALGQIWRRRGAAAFPPATRRRCRRRLPSRYHPPPPPPL
 LPCRVRSYLVGRRCRRRLPSCHHPSPSPPELLPGRHSRPLDVEDAATAAGPP
 RV"
 complement(38211..39233)
 /gene="B1074C08.8"
 complement(38211..39233)
 /gene="B1074C08.8"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAC00717.1"
 /db_xref="GI:21328138"
 /translation="MDWASLPADLLFSISSHLREPEDFVRPRVAVCPQWRAAVSHKEHA
 FQPMIARWLEDESENVLYSLTIKIVFVDFMKSRRIRASGSHLIADKOD
 DUSALINPLSGTALPLPKPFHDNGAHGWTGCVITVVLNNWMSNALWYRG
 GITMKGVAVPRKRLRVSHYLKLAAYQDMEMHLDLGDNEDSVLLQETKKVE
 LIGCWPRDSEVFKATPCHEWFSLYRVEQEIIPVHDIQNMVYVQSRDSCITRTYI
 PASRDFPALGSRNAPYLYLKQFDAGSYNLYFKCLASEVLTfVVKRLPEDNKLSDWEF
 MPSLKY"
 42943..43587
 /note="5' LTR"
 join(44015..44093,44667..44836)
 /gene="B1074C08.9"
 join(44015..44093,44667..44836)
 /gene="B1074C08.9"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAC00718.1"
 /db_xref="GI:21328139"
 /translation="WPPRRCLAAAPRVRPRASLPFCAAAEPGVADLSEGEDPDQVE

QGKSLFFHVDNCEIILFTNKIACNDHPTCDYVMP"
 join(47189..47443,47555..47728)
 /gene="B1074C08.10"
 Alignment Scores:
 Pred. No.: 7,07e-10 Length: 165701
 Score: 259.00 Matches: 104
 Percent Similarity: 39.34% Conservative: 75
 Best Local Similarity: 22.86% Mismatches: 162
 Query Match: 8.60% Indels: 114
 DB: 8 Gaps: 17
 US-09-913-878a-2_copy_710_1282 (1-573) x AP004357 (1-165701)
 Qy 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGly 20
 Db 112198 GGAATTGCGAAATATCCCGAGATCTTGCATTGCGAAGCTCCGAGCTGACA 112139
 Qy 21 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly 36
 Db 112138 GATNACCCGCACTGCTTATCAGATCAGGTTTCTGGCTTCAAGGGTGTCATAGTGC 112079
 Qy 37 TrpValIleAspValAspThrGlyAspGluAspTrpIleGluThrTyProSerGln 56
 Db 112078 TGG-----CAAGCACATGGTGATCGCACAGGGCTTTCTCTAGGCCAAGCATG 112031
 Qy 57 ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluAlaArgSerVal 76
 Db 112030 AGGAAGTTTGAGTCTAACCATTTG-----GTGTAGAGGTGGTCTCC--- 111989
 Qy 77 AlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArg 96
 Db 111988 TGGACAAAGTTCCAGCCAGGATTTCTTAATCGACAGATTATATATGCTATCTCA--- 111932
 Qy 97 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsnAspLeuGln 116
 Db 111931 -----CTGAATGTCCAGATCTCTATCTTTTGGCAATGCA 111896
 Qy 117 ArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpVal 136
 Db 111895 GAGACCATGCTTTCTAAC-----CTCAACAATATCTATCAGACAGAGATGTGCT 111845
 Qy 137 TyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 156
 Db 111844 TTTGAGGTTTTAAACAACCTTCATCTGTGT----- 111818
 Qy 157 GlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspPro 176
 Db 111817 -----GATGATGAAACACATGCGACATTGATGCTCAGTGTGGCTTTGAACCT 111770
 Qy 177 LysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThr 196
 Db 111769 AGAACTGAACACACACTTGAAGCAATGCTCTTGGCTATAAGTCTGCACAAATTCAGGAT 111710
 Qy 197 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 216
 Db 111709 CTTTGGAAAACAGAGGATTTTGGCCAAAGGAAGGTGGTGTGATGGGCTGCTGTGAT 111650
 Qy 217 PheTrpGlyValLeuGluGlu----- 224
 Db 111649 GAGCTTGGGGTCTTGAGCAAGGGCAGTGTCTTATTCTGGGCAACAGTTCATCATTTGAAT 111590
 Qy 225 GluValHisValGlyPheSerSerLysPheArgAspGluGluSerPheThrLeuLeu 244
 Db 111589 AGTTATTTTGTAAAGCATGGGTCAAGATTTTCATCAACAGATATAAACAACAGAGGTGATT 111530
 Qy 245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
 Db 111529 TTGGGT--ACTGTGGTAAATAGCAAGATCCCTGTCTTTCATCCAGGGGATGTCCGCATC 111473
 Qy 265 ValArgGlnAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 284
 Db 111472 CTTGAGCAGTGTGATGTCGCCGAACGTCACTGCTGTTGATTGTTGGTGTTCCTCCAG 111413


```

QY 238 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaAArgSerProAlaHis 257
   ::::::::::::::::::::
Db 2410 AAGACAGATCTGGAAGTAGTTAAAGGCTAT---GTAGCCATTTGTAAGATCCTTGCTT 2466
   ::::::::::::::::::::
QY 258 PheProSerAspIleGluArgValAlaValPheLysProGluLeuHisSerLeuLys 277
   ::::::::::::::::::::
Db 2467 CACCACAGGGATGAAGGATTTAGAACTGTTGATGTACCCAGCTGCATCATGTAT 2526
   ::::::::::::::::::::
QY 278 AspValIlePheSerThrLysGlyAspValProLeuAlaLysLeuSerGlyGly 297
   ::::::::::::::::::::
Db 2527 GACTGCTTATTTCCCTCAGAAAGGTGATAGCGCATACAAACGAAAGCTTCTGGCAGT 2586
   ::::::::::::::::::::
QY 298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheVal 317
   ::::::::::::::::::::
Db 2587 GACCTTGACGGAGACTGACTTTTGGCTTGGATCAGAACTCATC----- 2634
   ::::::::::::::::::::
QY 318 AsnAlaGluMetProLeuGluProAspLeuSerArgTyr-----LeuLysLysAspLys 335
   ::::::::::::::::::::
Db 2635 -----CCTCCCAACAGGAAAGCTATCGGGCCATCATTTATGATGCA 2676
   ::::::::::::::::::::
QY 336 ThrThrPheLysGluLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGluThr 355
   ::::::::::::::::::::
Db 2677 GCTGAAGAGAAGAGTTTA-----GGCGTGCTGTCAACACACAGGAC 2718
   ::::::::::::::::::::
QY 356 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375
   ::::::::::::::::::::
Db 2719 ATAATCGATTTCTTTCGAAGAAC-----TTGGCGAATGACGAGTGGGCACA 2766
   ::::::::::::::::::::
QY 376 CysThrAsn-----TyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 391
   ::::::::::::::::::::
Db 2767 ATTGCAATGCAACAGCTGCTGCTGATAGAGTGTAT-----GGAGCCCATG 2817
   ::::::::::::::::::::
QY 392 AsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 411
   ::::::::::::::::::::
Db 2818 GACGAAGATGTTGCTACTGACAGAACTAGTGCACCTGCGAGTTGATTTCCAAAGACA 2877
   ::::::::::::::::::::
QY 412 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgGluLeuLeuGlyGlyAla 431
   ::::::::::::::::::::
Db 2878 GGGAAATTTGTGTCATGCCCC-----TTCCACCTTAAACCAACCACTTACCCAGATTTC 2931
   ::::::::::::::::::::
QY 432 LeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeuGlyArg----- 447
   ::::::::::::::::::::
Db 2932 ATGGGAAAGAGACTACCAAACTTACAAGTGCACAAATCTTGGTCGGCTTTACAGA 2991
   ::::::::::::::::::::
QY 448 -----GlyGluPro 450
   ::::::::::::::::::::
Db 2992 CGGTAAAGAGGTTTATGATGAAGATGCAGAAAGTCTCTCAGAAAGAACACAGACCA 3051
   ::::::::::::::::::::
QY 451 ThrHisIle-----IleAspTyrLeuLysPheSerIleAlaArgPro--- 464
   ::::::::::::::::::::
Db 3052 AGTGCCATCCCTATGACCTGTTCTTGAATACCGGANTTTGAAGATTTGATCCCTGAG 3111
   ::::::::::::::::::::
QY 465 -----AlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 478
   ::::::::::::::::::::
Db 3112 GCATGGGTGCACAAATGTTGTACACGGGCAACTCATTTGCTTCTTGGGCAATACAA 3171
   ::::::::::::::::::::
QY 479 AlaAlaLysAspThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
   ::::::::::::::::::::
Db 3172 GTCCAGAAAGAGAGAGATTTGACGGGTACATCTGTCC-----ATGCCCAAA 3222
   ::::::::::::::::::::
QY 498 TyrThrPhePheLysGluLeuSerAspLysSerArgSerSerAlaLeuPheThrThr 517
   ::::::::::::::::::::
Db 3223 TACACAAGCAAGAAACAGGCAAGCACTGAAAGAAAGA----- 3258
   ::::::::::::::::::::
QY 518 LeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGlu 537
   ::::::::::::::::::::
Db 3259 CTGAACGATCTTATTAATTCCTTTAAAGAGAGTTC---AGGAAAGTATTTCAGGAAACA 3315
   ::::::::::::::::::::
QY 538 MetArgAspSerLysAspProTyrProValArgValAsnGlnValTyrGluLys----- 555
   ::::::::::::::::::::
Db 3316 ATCCCTGACCATGAAATCTCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCT 3375
   ::::::::::::::::::::
QY 556 -----TrpCysAlaIleThr-----ProGluAlaMetAspLysSer 567

```

Db 3376 TCAGCTTGATCATGATAACTTACCATCCGAGTGGGTGAAGAGTCT 3423

RESULT 14

AF443073 3524 bp DNA linear PLN 03-DRC-2001
Phomopsis sp. CMW 5588 putative RNA-dependent RNA polymerase
(rdp-1) gene, partial cds.

ACCESSION AF443073 GI:17227101

VERSION AF443073.1

KEYWORDS

SOURCE

ORGANISM

Phomopsis sp. CMW 5588

Phomopsis sp. CMW 5588

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariomycetidae; Diaporthales; Valsaceae; mitosporic Valsaceae;

Phomopsis.

REFERENCE 1 (bases 1 to 3524)

AUTHORS Preissig,O. and Wingfield,M.J.

TITLE Direct Submission

JOURNAL Submitted (04-NOV-2001) Forestry and Agricultural Biotechnology

Institute, University of Pretoria, 72 Lunnun Street, Pretoria 0002,

South Africa

FEATURES

Location/Qualifiers

1..3524

/organism="Phomopsis sp. CMW 5588"

/mol_type="genomic DNA"

/isolates="CMW 5588"

/db_xref="taxon:179291"

/country="South Africa"

<1..>3424

/genes="rdp-1"

join(<1..2092,2148..>3424)

/genes="rdp-1"

/products="putative RNA-dependent RNA polymerase"

join(<1..2092,2148..>3424)

/genes="rdp-1"

/note="RDP-1"

/codon_start=1

/product="putative RNA-dependent RNA polymerase"

/protein_id="AAL38011.1"

/db_xref="GI:17227102"

/translation="HKLHLPASLVKPVQSQVFGRTTTFYFINSFSLSWESVRT

YLAADVNVKTVQVNRFEIVTPGSNAWRLDLSAEMTLELLSSNTYHLDFOVRYOLEV

CTSHGLDEVSVNAQFLDKNSFADARMMLLEGVAESIQHLEPMKIFDDPKVLHYM

PHARVPYATLVRRAVITFTTYFKTPCVELTNRLKYSGLNDLFLRVQFTDITFG

KIPSQDSNKNLNIYIRVRHWQNGISIGDRHYRFLAFNSQFRENAGFFCETDHT

CDISIREWMDGFRHISVGFAPARMGQCFPTTQVNGISIPNKKIDDDIERQIGDHTW

FTDVGKISNFFAKLTASERDLFETPSCFMRIQGGKGLVVMVDVPPSEVHRPSQE

KFVAVNVLIIKTFATSHATFNKQIVPVLVALGVDSVFSVMDDELEXYDEALDS

MKAGELLRSQVDENOTTLTMAEMVDTFMDKEPFLWTLRLKWLKMLKELKHFATISV

KKSAMVGVVDIGVLGRHSQATGRGYSIESLPQIFLOVPIEGSDGSRSTNYEVIT

GLCVGRNPSLHPGDVRYVAVDVPELHLKRVVFPKTDGDRDIPSMCSGSGMDGDDY

FVTWDERLPTEDHPPLNHDAGSSTLDKPADVTIEDVTRFFAQYKNDLSGLRIATAH

FAQDQLSGGVKNPKICIELAKLHSMADVTKSGRFAVMKGRHLQPKRMPHMEKEKSN

YRSYALGKIYDKIETIEPAAAYEMPFDARILSRVQMGKADTLAKASKIKATDIANRR

LMQHEAPATFEPIEFTSLRPSVSGSLQLENVGRLEALKLFRACMEAVTQNI

QTEGAPASVVNLEKLDLFVAAMTVTHNDVRAALRERMPKPNEGSGSEIQMPLI

SPFMLPHRELARVALGRGDVRLPQKSVNLNGESQOTDRDRAVEDADFEIPGDTAL

PDALILEVTQQAQVGVDSQKSGYTTTGRGSGADGIEGDDCVRTSTGQVVHFGMLAL

FTTGEVQCQQRSMPTDISPAGTGLQSDASVSNNLSLVADDSTGSSCCPASPRSERE

VVEEDSDIETVEVDGAGIEVEDEALALARKLGI"

ORIGIN

Alignment Scores:

Pred. No.: 1..16e-11 Length: 3524
Score: 254.00 Matches: 140
Percent Similarity: 39.02% Conservative: 84
Best Local Similarity: 24.39% Mismatches: 216
Query Match: 8.44% Indels: 134
DB: 8 Gaps: 25

US-09-913-878A-2_COPY_710_1282 (1-573) x AF443073 (1-3524)

QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgGleAspValLeuGlyLeuGly 20
 DB 1012 GGAGTGGGCGAGATATCAACTTTTTCGAAAACATGATCGAAGGCGGATTTGCCA 1071
 QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
 DB 1072 GAGAGCCCTTCATGCTTCAGATGAGATAGCGCGCGCAAGGGCGTCTCGTGGTGG 1131
 QY 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyProSerGlnArgLysTrpGlu 60
 DB 1132 CCTGACGTTCCCAAGCGAA-----GTTACGTCGACCTTCTCAAGAGAAATTCAG 1185
 QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
 DB 1186 GCTGTGTAT-----AACGTCCTCGAGATCATCAAGACCTTTGCA-----TTT 1227
 QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaAspLys 100
 DB 1228 TCCACGCCCACTTCAATAAGCAGGTATACCTGCTAGTCGCCCTTGGTGTGGAC--- 1284
 QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
 DB 1285 -----AATTCGCTTCGTTAGCATGCTCGATGATGAGCTCAGAGACTACGAC 1332
 QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyGluSerTy 140
 DB 1333 GAA-----GCTTTC-----GCAGAT 1347
 QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
 DB 1348 TCCATGAAGCTGTGCACTCTTCGATCCCAAGTC-----GAT 1386
 QY 161 SerGlnGluThrLeuAsnPheLeu---MetAsnSerGlyPheAspProLysGln 179
 DB 1387 GAAACCAACGACATACCATGCGCAGAGATGGTCGACACCTTCATGAGCTCCAAGGAG 1446
 QY 180 LysTyLeuGlnAspIleAlaTrpAspLeuGlnLysArg---LysCys-----Asp 195
 DB 1447 CATTCCTT-----TGGACCTTGCTTCCTTGTGGAGTGTGATATTGAG 1494
 QY 196 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyIleTyMetIleAla 215
 DB 1495 AGATTGAAGCACAAATTCGCATCATGCTCAAGAAAGCGCCATGCTTCGGGTGTC 1554
 QY 216 AspPheTrpGlyValLeuGluGluAsnGluValHsVal-----GlyPheSerSer 232
 DB 1555 GACGAGATTGGTGTCTGAGGGGTCACTCGCAAGCCACTGAAGGCGAGGATACATAGC 1614
 QY 233 -LysPheArgAspGluGluGluSerPhe----- 241
 DB 1615 ATCGAGTCGCTCCCAAAATCTCTTCAAGTGCCTATTGAGGCGAGCGATGGGAGGAGT 1674
 QY 242 -----ThrLeuLeuSerAspCys-AspVal-LeuValAlaArgSerProAlaHisPhe 258
 DB 1675 ACCACCAACTATGAAGTCACTACGCGATTTGCGTGTGGCGCAATCCATCACTCCAT 1734
 QY 259 ProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAsp 278
 DB 1735 CTGGAGATGACGAGTGTGTCAGGAGTAGAGCGTGCCTGAGCTTCGATCTGAAGAAC 1794
 QY 279 ValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAsp 298
 DB 1795 GTGCTGCTTCCCAAGACTGCGCAGAGACATTCGAGCATGTGTGTGGAGGTGAC 1854
 QY 299 TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAsp----- 314
 DB 1855 ATGGACGGGATGACTACTTGTCTATGGATGAGAGACTCATTCACACTGAGTGGGAC 1914
 QY 315 -----GlyPheValAsnAlaGluMetProLeuGlu----- 324
 DB 1915 CATCCACCCCTGAACACGATCGGGATCATCCACTTGGCAAGACCGCGCAGAGCTACA 1974
 QY 325 ---ProAspLeuSerArgTyLeuLysLysAspLysThrThrPheLysGlnLeuMetAla 343

DB 1975 ATTGAGGAGCTCACCAGATTC-----TTTGCACAGTATATGAAG 2013
 QY 344 SerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyAspMetIleGlnLysSer 363
 DB 2014 AATGACTCCCTGGCGCGTATCGCC-----ACT 2040
 QY 364 PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyLysGluArgLeu 383
 DB 2041 GCGCACTTTGCCAGGAGCATCACTGTCAGGGGCGTCAAGAACCCAAAA----- 2091
 QY 384 Cys-----TyrIleAsn-AsnSerValSerAsnLysProAlaI 396
 DB 2092 TGTAAAGCATGCCCTATCAGCTTATTCTTGATGAACAACTCTGACAGCTCGAATAGGCAT 2151
 QY 396 eIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAs 416
 DB 2152 CGAACTCGCAAGCTCACTCCATGCTGCTCGACTATATTAAGTCTGGC----- 2200
 QY 416 nGluAlaSerTrpAlaGlnLeuArgGluLeuLeuGlyGlyAlaLeuSerLeuProAs 436
 DB 2201 -----AGACCTGCGCTCATGAAGAGACACCTC-----CA 2229
 QY 436 pProMetTyLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTy 456
 DB 2230 ACCTCGA---AATGCGCGCATTTGATGAAGAGAGAGAGAGTCT-----AACTA 2277
 QY 456 rLeuLysPheSerIleAlaArgProAlaIleAspLys---GluLeuGluAlaPheHisAs 475
 DB 2278 TCGTCTTACTCGCGCTCGCAAGATCATGACCGGATCAAGATCAAGATTCATTCATGC 2337
 QY 475 nAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAl 495
 DB 2338 GGATACGAAATGCCCTTTGACGCCGCG-----ATCCTCTC 2373
 QY 495 aserTyTyThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPh 515
 DB 2374 GCGCTACGAGTACAGCAGATACCTGCGCAAGGCGCAAAATCAAGCCACATATGA 2433
 QY 515 eThrThrLeuLysAsnArgIleGlyGluValGlu 526
 DB 2434 CATCGCATCGCGCGCTCATGGCAACACGAG 2467
 RESULT 15
 AC105931/c 132716 bp DNA linear HTG 11-JAN-2002
 LOCUS Magnaporthe grisea chromosome 7 clone 21J15, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 31 unordered pieces.
 AC105931
 AC105931.1 GI:18129404
 VERSION HTG; HTGS PHASE1.
 KEYWORDS Magnaporthe grisea
 SOURCE Magnaporthe grisea
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
 REFERENCE 1 (bases 1 to 132716)
 AUTHORS Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr, Brown,D.E., Taro,A. and
 Papalas,J.
 TITLE Magnaporthe Grisea chromosome 7
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 132716)
 AUTHORS Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr and Brown,D.E.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2002) Plant Pathology - Fungal Genomics
 Laboratory, North carolina State University, 840 Main Campus Drive,
 Raleigh, NC 27606, USA
 COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will

| | | |
|--------------------------------|---------------------------------------|----------------|
| * be preserved. | 981: contig of 981 bp in length | |
| * | 1081: gap of unknown length | |
| * | 1082: contig of 711 bp in length | |
| * | 1793: gap of unknown length | |
| * | 1893: contig of 601 bp in length | |
| * | 2494: gap of unknown length | |
| * | 2594: contig of 715 bp in length | |
| * | 3309: gap of unknown length | |
| * | 3409: contig of 838 bp in length | |
| * | 4247: gap of unknown length | |
| * | 4347: contig of 266 bp in length | |
| * | 4613: gap of unknown length | |
| * | 4713: contig of 1418 bp in length | |
| * | 6131: gap of unknown length | |
| * | 6231: contig of 743 bp in length | |
| * | 6374: gap of unknown length | |
| * | 7074: contig of 1626 bp in length | |
| * | 8700: gap of unknown length | |
| * | 8800: contig of 2036 bp in length | |
| * | 10896: gap of unknown length | |
| * | 10996: contig of 1978 bp in length | |
| * | 12974: gap of unknown length | |
| * | 13074: contig of 2338 bp in length | |
| * | 15412: gap of unknown length | |
| * | 15511: contig of 3560 bp in length | |
| * | 19072: gap of unknown length | |
| * | 19171: contig of 2690 bp in length | |
| * | 21861: gap of unknown length | |
| * | 21961: contig of 2549 bp in length | |
| * | 24510: gap of unknown length | |
| * | 24511: gap of unknown length | |
| * | 27958: contig of 3348 bp in length | |
| * | 28058: gap of unknown length | |
| * | 32486: contig of 4428 bp in length | |
| * | 32586: gap of unknown length | |
| * | 36925: contig of 4339 bp in length | |
| * | 37025: gap of unknown length | |
| * | 43846: contig of 6821 bp in length | |
| * | 43946: gap of unknown length | |
| * | 48950: contig of 5004 bp in length | |
| * | 49050: gap of unknown length | |
| * | 54904: contig of 5854 bp in length | |
| * | 55004: gap of unknown length | |
| * | 60187: contig of 5183 bp in length | |
| * | 60287: gap of unknown length | |
| * | 65586: contig of 5299 bp in length | |
| * | 65686: gap of unknown length | |
| * | 71720: contig of 6034 bp in length | |
| * | 71820: gap of unknown length | |
| * | 80013: contig of 8193 bp in length | |
| * | 80113: gap of unknown length | |
| * | 88629: contig of 8516 bp in length | |
| * | 88729: gap of unknown length | |
| * | 96184: contig of 7455 bp in length | |
| * | 96284: gap of unknown length | |
| * | 102718: contig of 6434 bp in length | |
| * | 102818: gap of unknown length | |
| * | 112425: contig of 9607 bp in length | |
| * | 112525: gap of unknown length | |
| * | 120940: contig of 8415 bp in length | |
| * | 121040: gap of unknown length | |
| * | 132715: contig of 11676 bp in length. | |
| Location/Qualifiers | | |
| 1. 132716 | | |
| /organism="Magnaporthe grisea" | | |
| /mol_type="genomic DNA" | | |
| /db_xref="taxon:148305" | | |
| /chromosome="7" | | |
| /clone="21J15" | | |
| FEATURES | source | |
| ORIGIN | | |
| Alignment Scores: | | |
| Pred. No.: | 1.41e-03 | Length: 132716 |

| | | | |
|--|--------|---|-----|
| Score: | 254.00 | Matches: | 118 |
| Percent Similarity: | 37.83% | Conservative: | 56 |
| Best Local Similarity: | 25.65% | Mismatches: | 158 |
| Query Match: | 8.44% | Indels: | 128 |
| DB: | 2 | Gaps: | 21 |
| US-09-913-878A-2_COPY_710_1282 (1-573) x AC105931 (1-132716) | | | |
| Qy | 1 | GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20 | |
| Db | 108509 | GGCGTTGGCAAAATCTAGCTGTGGCTGTGTCTATTCAGAGATCTCCAGAGCTGGAC 108450 | |
| Qy | 21 | AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40 | |
| Db | 108449 | TATATCCCTTCGGCGTTCAGTCCGTATGCGGCGCTGTAAAGGTCTTCTAGTGGTCTG 108390 | |
| Qy | 41 | ValAspAspThrGlyAspGluAspTrpIleGluThrTrpProSerGlnArgLysTrpGlu 60 | |
| Db | 108389 | GAAGACGTGAAGGCCAAGAG-----ATTACATACGTCCTTCTCAGCTCAAGTTTGG 108336 | |
| Qy | 61 | CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80 | |
| Db | 108335 | ACGGAC-----CATAGTCGCTTGAAGTCAATTCGC-----TGTCTCACTTC 108294 | |
| Qy | 81 | LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys 100 | |
| Db | 108293 | TCTGTGGCGTCTCTCAACGACAAACAATCTCGTCTTG----- 108255 | |
| Qy | 101 | ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeu---GlnArgGlnPhe 119 | |
| Db | 108254 | -----ACTCCCTCGGAGTCAAA-----AATGAGGTGTTGAAGATATGTTG 108213 | |
| Qy | 120 | SerGluGlnLysHisAlaLeuAsnArgProValGlu----- 131 | |
| Db | 108212 | GCAGGCGAGTGGCAAAATACACAGGCGATGAAGATCGCAGCAGCATGAAGCTC 108153 | |
| Qy | 132 | PheArgGlnTrpValTyrGluSerTyrSerArgAlaThrArgValSerHisGlyArg 151 | |
| Db | 108152 | CTGACACAGTACGTTGACGAA----- 108132 | |
| Qy | 152 | ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 171 | |
| Db | 108131 | -----AATCAATGACCAACACATTTGGCCGCGCTGTGAAG 108096 | |
| Qy | 172 | SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIle-----Ala 186 | |
| Db | 108095 | GCTGGCTTATGGAGTCGCATGAGCCTTTGTCTGCACCATGCTGCGCCTATGAGGACT 108036 | |
| Qy | 187 | TrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGly 206 | |
| Db | 108035 | TGGTCAGTTAAG-----ACTCTCAGGAAAGCAAGACTCGTCGTGAG 107991 | |
| Qy | 207 | ArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluVal 226 | |
| Db | 107990 | AAGGTGTCANTCGTTCTTGGCTGTGTGCGAAGAACCAAGACCTTACATGGCCACTTCAAC 107931 | |
| Qy | 227 | HisValGlyPheSerSerLysPheArgAspGlu-----GluGlu 239 | |
| Db | 107930 | CATATCCCGCCAGTCCCAACAGAGGCCCGGCTCAATGGCAATATGAAGCTGCCTGAG 107871 | |
| Qy | 240 | SerPheThrLeuLeuSerAsp-----Cys 247 | |
| Db | 107870 | ATATTCTTCAGTCCCTGATGGCCCTAGCGCTGACTATATAAGTCATACCGGACTTTGC 107811 | |
| Qy | 248 | AspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAla 267 | |
| Db | 107810 | -----CTTGTGGCCGCAACCCCTCTTACCCCTGGAGACATCCCGGTGCTCCAGGCA 107757 | |
| Qy | 268 | ValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAsp 287 | |
| Db | 107756 | ATTGATGTCCTCAATTGGCGACACATCAGGAGCGTCTGTGTTTGTGTTTGGCCCAACAGGAGAT 107697 | |
| Qy | 288 | ValProLeuAlaLysLysLeuSerGlyLysAspTyrAspGlyAspMetAlaTrpValCys 307 | |


```
Db 107696 AGAGACATTCCTGGCANGTGCTCTGGAGGAGACTTGGATGGTGACGATTTCTGCTCTTG 107637
QY 308 TrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeu 327
Db 107636 TGGGATGACGGGCTT-----CTCCCTTTA---CCTTCGCAG 107604
QY 328 SerArgTyrLeuLysLysAspLysThrPheLysGlnLeuMetAlaSerHisGlyThr 347
Db 107603 CGCAATTATTAAGCCG-----GCAAGCTATCATCCC 107574
QY 348 GlySerAlaAla-----LysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 364
Db 107573 CCTCCTGCCCGAGAGCTTGACCGGGAAGTTACGGCAGATGATCTTGGCAG-----TTT 107520
QY 365 HisPheAlaLeuGlnProAsnPhe-----LeuGlyMet 375
Db 107519 TTTGTGACATACATGAAGAACCTTTCACTTCCAAGGATTGCTCATGCTCACCTGGCGCAG 107460
QY 376 CysThrAsnTyr-----LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLys 393
Db 107459 GCAGACAATATGGCGCAAAAGCAGAGAGTGT----- 107427
QY 394 ProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIle 413
Db 107426 -----CTTGAAATGGCCTATTGTCACCTCGTTGGCGGTTGATTATCCCAAAACGGGAGTG 107373
```

Search completed: March 31, 2004, 11:00:02
Job time : 3775.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 02:00:42 ; Search time 340.609 Seconds
(without alignments)
7146.671 Million cell updates/sec

Title: US-09-913-878A-2_COPY_710_1282

Perfect score: 3010

Sequence: 1 GVGRMSRVAKIRIDVLGLG.....EKWCAITPEAMDKSGANYDS 573

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_epool/US09913878/runat_29032004_082007_8357/app.query.fasta_1.2254
-DB=N_Geneseq_29Jan04 -QWTF=fastap -SUFFIX=eng -MINMATCH=0.1 -LOOPTCL=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878 -CGN 1 1 797 @runat_29032004_082007_8357 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Match | Length | ID | Description |
|--------|-------|-------|-------|--------|----------|-------------|
| 1 | 3010 | 100.0 | 8045 | 3 | AAA65171 | Neurospora |
| 2 | 341 | 11.3 | 615 | 7 | ABZ54843 | Aspergillus |
| 3 | 262 | 8.7 | 3731 | 4 | AA517845 | Tomato RN |
| 4 | 262 | 8.7 | 3731 | 4 | AA517837 | Tomato RN |
| 5 | 262 | 8.7 | 3731 | 4 | AA517837 | Tomato RN |
| 6 | 258 | 8.6 | 3591 | 4 | AAH77704 | Nucleotide |
| 7 | 255 | 8.5 | 2816 | 3 | AA63741 | Rice RNA |
| 8 | 250.5 | 8.3 | 3807 | 3 | AA63742 | Soybean R |

| | | | | | | |
|----|-------|-----|--------|---|-------------|--------------------|
| 9 | 240.5 | 8.0 | 3737 | 3 | AA63739 | Maize RNA |
| 10 | 238.5 | 7.9 | 3901 | 3 | AA63740 | Maize RNA |
| 11 | 220.5 | 7.3 | 6863 | 4 | AAH77703 | Nucleotide |
| 12 | 218 | 7.2 | 1281 | 3 | AA63738 | Maize RNA |
| 13 | 148.5 | 4.9 | 4818 | 4 | ABL27117 | Drosophila |
| 14 | 148 | 4.9 | 282 | 6 | ABL71773 | Corn tass |
| 15 | 132 | 4.4 | 274 | 6 | ABL71722 | Corn tass |
| 16 | 126.5 | 4.2 | 3984 | 4 | AAH45252 | Human pol |
| 17 | 126.5 | 4.2 | 7450 | 4 | AAH45260 | Human pol |
| 18 | 125 | 4.2 | 110000 | 6 | ABA90521_11 | Concinnation (12 o |
| 19 | 124 | 4.1 | 274 | 7 | ABX87270 | Corn ear- |
| 20 | 124 | 4.1 | 1701 | 5 | AA577554 | DNA encod |
| 21 | 123.5 | 4.1 | 16122 | 4 | AA522908 | DNA encod |
| 22 | 123.5 | 4.1 | 16341 | 4 | AA523002 | DNA encod |
| 23 | 120 | 4.0 | 9274 | 6 | ABK13224 | DNA encod |
| 24 | 118.5 | 3.9 | 5841 | 6 | ABZ23259 | Candida a |
| 25 | 117.5 | 3.9 | 11241 | 6 | ABK82005 | CDNA encod |
| 26 | 117.5 | 3.9 | 11443 | 6 | ABK82002 | DNA encod |
| 27 | 117.5 | 3.9 | 12323 | 1 | AA903338 | Sequence |
| 28 | 117.5 | 3.9 | 13957 | 6 | ABK81959 | cdNA encod |
| 29 | 117.5 | 3.9 | 13957 | 6 | ABT10904 | Human bre |
| 30 | 117.5 | 3.9 | 13957 | 6 | ABN95786 | Gene #228 |
| 31 | 117.5 | 3.9 | 13957 | 6 | ABN95786 | Gene #228 |
| 32 | 117.5 | 3.9 | 13977 | 6 | ABN95786 | Human bon |
| 33 | 116.5 | 3.9 | 12154 | 9 | ABN95786 | Human bon |
| 34 | 116.5 | 3.9 | 8114 | 4 | ABL03787 | Drosophila |
| 35 | 116 | 3.9 | 6006 | 4 | AAH41086 | Human cal |
| 36 | 115.5 | 3.8 | 3588 | 7 | ACA28683 | Prokaryot |
| 37 | 115 | 3.8 | 9630 | 5 | AA587490 | DNA encod |
| 38 | 114.5 | 3.8 | 8370 | 5 | AA513652 | CDNA sequ |
| 39 | 114.5 | 3.8 | 12461 | 5 | AA513653 | Genomic D |
| 40 | 114 | 3.8 | 110000 | 6 | ABA90521_00 | Genomic s |
| 41 | 113.5 | 3.8 | 11058 | 6 | AA372229 | Human dys |
| 42 | 113 | 3.8 | 2793 | 7 | ACA43117 | Prokaryot |
| 43 | 113 | 3.8 | 6893 | 2 | AA32182 | Human coa |
| 44 | 113 | 3.8 | 8494 | 2 | AAV52296 | Streptoco |
| 45 | 113 | 3.8 | 110000 | 7 | ABN56454_06 | Continuation (7 of |

ALIGNMENTS

RESULT 1

AAA65171

ID AAA65171 standard; DNA; 8045 BP.

XX

AC AAA65171;

XX

DT 16-NOV-2000 (first entry)

XX

DE Neurospora crassa qde-1 gene.

XX

KW Gene silencing; quelling deficient; qde-1; al-1; ds.

XX

OS Neurospora crassa.

XX

PH Location/Qualifiers

FT CDS

FT 2447..6655

FT /tag=a

FT /product="qde-1"

XX

PN WO200050581-A2.

XX

XX

PD 31-AUG-2000.

XX

PF 16-FEB-2000; 2000WO-IT000048.

XX

PR 22-FEB-1999; 99IT-RM000117.

XX

PA (UYRO-) UNIV ROMA LA SAPIENZA.

XX

PI Macino G, Cogoni C;

XX

DR WPI; 2000-579171/54.

DR P-PSDB; AAB13956.
XX Novel polynucleotide encoding a polypeptide which has a silencing
PT activity and comprising a RNA-dependent RNA polymerase domain.
XX
XX
PS Claim 1; Page 31-43; 48pp; English.
XX

CC The present sequence is the *Neurospora crassa* qde-1 gene. This gene has
CC silencing activity. The qde-1 gene was isolated by mutational analysis of
CC an al-1 transgenic strain. This strain had an albino phenotype resulting
CC from post-transcriptional silencing of the endogenous al-1 gene.
CC Reversion of this phenotype indicated a mutation in a silencing gene. The
CC silencing gene, qde-1, could then be isolated. Modulation of qde-1
CC expression may be used to inactivate genes and to silence suppression of
CC genes

SQ Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 U; 0 Other;

Alignment Scores:

Prod. No.: 8 23e-283 Length: 8045
Score: 3010.00 Matches: 573
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-913-878a-2_COPY_710_1282 (1-573) x AAB65171 (1-8045)

QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
DB 4574 GGTGTAGGCGCATGTCGCGAAGCGTGGCAAGAGATACGGGATGTTCTCGGTTGGGT 4633
QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
DB 4634 GATGTGCGCTCTGCTGTGCAGAGGCGGTTGGTTCGGCCCAAGGGATGTGGGTATCGAC 4693
QY 41 ValAspThrGlyAspGluAspTrpIleGluThrTrpProSerGlnArgTrpGlu 60
DB 4694 GTTGACGACACAGCGCATGAGGATTCGATCGACATACCCGTCCAGCGCAAGTGGGAA 4753
QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
DB 4754 TGGACTTCGTGATTAACATCAACGTACCTCGAAGTCCGAGCGTGGCTTCTGAATG 4813
QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
DB 4814 AAGTCAGTGTCTCAACCTACAGCTGTTTACCTGTCTCTGGAAGATAGAGCCAGGACAA 4873
QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
DB 4874 GTGAAGATGCCAGGCAATCGGTGACCGCTTATCAACGATTTGCAACGACAGTTCAGC 4933
QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTrpGluSerTrp 140
DB 4934 GAGCAAAAGCATGCTTTGAATGCCCGCAGTGGAAATTTCCGCAATGGTTTACGAGAGTTAT 4993
QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
DB 4994 TCCAGTCGCGCAACTCGAGTCAGCCACGCGCGTGTGCTTTTCTGTGGGTACTCTGAC 5053
QY 161 SerGlnGluGlnThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
DB 5054 AGTCAAGAGGAGACACTGAATCTTGTATGAACAGTGGGTTTCGATCCCAAGAGCAAG 5113
QY 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
DB 5114 TACTTCAAGACATCCCTGGGATCTTCAAAAGCGGAAATGTGCACGTTGAAGTCCAAG 5173
QY 201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
DB 5174 CTGAACATCGGTGCTCGATCAGCATACATTTACATGATTTGCCGATTTCTGGGGTGTG 5233
QY 221 LeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluSer 240

DB 5234 CTTGAGAAAATAGGTTTCATGTCGGATTCCTCCTCAAGATTCAGAGCCAGGAGGAGTCT 5293
QY 241 PheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 260
DB 5294 TTTACACTCTCTATCGGACTGTGATGCTCTGCTGGCGGATCCCGAGCCCAATTCCTAGT 5353
QY 261 AspileGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
DB 5354 GATATCAACAGGGTTCGAGCAGTCTTCAAGCCAGAGTCCACAGTCTCAAGGATGTAATC 5413
QY 281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTrpAsp 300
DB 5414 ATCTTCTTACTAAGAGGATGATACCGCTTGTCTAAGAGCTATCTGGTGGAGACTACGAC 5473
QY 301 GlyAspMetAlaTrpValCysTrpAspProGluLeuValAspGlyPheValAsnAlaGlu 320
DB 5474 GGCATATGGCTGGGTCTGCTGGATCCGAGATCGTCATGGTTCTGTCATATCGGAA 5533
QY 321 MetProLeuGlnProAspLeuSerArgTrpLeuLysLysAspLysThrTrpPheLysGln 340
DB 5534 ATGCCTCTGGAGCCCGACCTGTCTAGGTACTTAAAGAGGACAAAGAGCTTTCACAA 5593
QY 341 LeuMetAlaSerHisGlyThrGlySerAlaLysGluGlnThrTrpTrpAspMetIle 360
DB 5594 CTTATGGCTTCACAGCGCAGGCTCAGCGGCCCAAGAGCAGACTATACGATATGATC 5653
QY 361 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTrpLys 380
DB 5654 CAGAAGAGCTTCCATTTTCGCCCTGCAGCCCACTCTTGGCGATGTCACATACTACAA 5713
QY 381 GluArgLeuCysTrpIleAsnAsnSerValSerAsnLysProAlaIleLeuSerSer 400
DB 5714 GAAAGGCTCTGTATACATCAACAATAGTGTCTTAAACAGCGGCCCATCATCTTAGTTCA 5773
QY 401 LeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrp 420
DB 5774 CTGTGTGGAAACCTCGTCGATCAGACCAAGAGTATTGTCTTTAACGAAAGCAAGCTGG 5833
QY 421 AlaGlnLeuArgArgGluLeuLeuGlyValAlaLeuSerLeuProAspProMetTrpLys 440
DB 5834 GCTCAATTTGGTAGGAGACTGCTTGGCGGTGCATTTGTCTCTTCCCTGACCCCAATGTACA 5893
QY 441 SerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTrpLeuLysPheSer 460
DB 5894 AGCCACAGTTGGTTCGGCGCGGAGAGCTTACCCACATTTATTGACTACTCTGAAATTCCTC 5953
QY 461 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAla 480
DB 5954 ATCCCGAGGCTCGGATTCAGAGAACTGGAAGCCCTTCCCAATGCCATGCAAGAGCGGCC 6013
QY 481 LysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTrpTrpPhe 500
DB 6014 AAGGATACAGAGACGCGCTCACTTTTGGGATCCGATCTCGCTTCTCTACTACAGCTTC 6073
QY 501 PheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrTrpLeuLysAsn 520
DB 6074 TTCAAGAGATTAAGCAAGTCCGATCGCTCGCAGCTGTCTATTACGACTCTGAAGAAC 6133
QY 521 ArgIleGlyValGluLysGlyTrpGlyArgLeuValLysAsnLysGluMetArgAsp 540
DB 6134 CGTATCGCGAGTTCGAGAAAGAAATATGCGAGTGTGGTCAAAAACAAGAGATGAGAGAC 6193
QY 541 SerLysAspProTrpProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThr 560
DB 6194 AGCAAGGACCCCTACCTGTCGCGCTCAACAGGTTTATGAAAATGGTGGCCATCACG 6253
QY 561 ProGluAlaMetAspLysSerGlyAlaAsnTrpAspSer 573
DB 6254 CCTGAGCGGATGACAAATCCGAGCAAAATTAATGATCT 6292

RESULT 2
ABZ54843

ID AB254843 standard; cDNA; 615 BP.
 XX AC AB254843;
 XX DT 28-MAR-2003 (first entry)
 XX DE Aspergillus oryzae polynucleotide SEQ ID NO 3956.
 XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
 XX KW expressed sequence tag; gene; ss.
 XX OS Aspergillus oryzae.
 XX PN WO200279476-A1.
 XX PD 10-OCT-2002.
 XX PF 22-MAR-2002; 2002WO-IB000890.
 XX PR 30-MAR-2001; 2001JP-00098371.
 XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX PA (NARE-) NAT RES INST BREWING.
 XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.
 XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
 XX PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
 XX DR WPI; 2003-046817/04.
 XX CC Detection of expression of specific Aspergillus genes for monitoring the
 XX PT fermentation and growth conditions of the fungus, using DNA probes.
 XX PS Claim 1; SEQ ID NO 3956; 48bp + Sequence Listing; Japanese.
 XX CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (AB250888-AB256893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridising
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of Aspergillus oryzae which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences
 XX SQ Sequence 615 BP; 151 A; 148 C; 162 G; 154 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8,498-24 Length: 615
 Score: 341.00 Matches: 89
 Percent Similarity: 54.75% Conservative: 32
 Best Local Similarity: 40.27% Mismatches: 80
 Query Match: 11.33% Indels: 21
 DB: 7 Gaps: 6

US-09-913-878A-2_COPY_710_1282 (1-573) x AB254843 (1-615)

QY 226 ValHisValGlyPheSerSerLysPheArg-AspGluGluSerPhe-----ThrLe 243
 Db 1 GTTCACTTTGGCTTTCTCCAGTAACCTGGCTGATCGTGAAGCCCAATCGAAGCAATCT 60
 QY 243 uLeuSerAspCysAspValLeuAlaAlaArgSerProAlaHisPheProSerAspIleG1 263
 Db 61 ACTCGACGGTATGGACGTTCTTTGTAGCTCGCTCCCGCACATGTCGGCTGATATCCA 120
 QY 263 nArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIlellePheSe 283
 Db 121 CGGACGAGAGAGCTGTGTGAGAGCCAGAGCTCCGCATTTTCAAGATGTTATCGTTTCC 180
 QY 283 tThrLysGlyAspValProLeuAlaLysLeuSerGlyGlyAspTyrAspGlyAspMe 303

Db 181 TACTGTTGGACAAACGCCACTGGCGCACATGTTATCCGGTGGTGATTACGAGGTATAC 240
 QY 303 tAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet---Pr 322
 Db 241 AACGTGGGTTTGCTGGGACCAAAATATTGTCAGAAAGTTTCGTAATTCGACCTATCCAC 300
 QY 322 oLeuGluProAspLeuSerArgTyr---LeuLysLysAspLysThrThrPheLysGlnLe 341
 Db 301 CATGGATACCCCTGCGGAGCACTTTGGTCTTGAGGAACATAACGTGCCCAATGAAGATAT 360
 QY 341 uMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleG1 361
 Db 361 T-----GACTCATGGGACGAGTTTCTTCA 384
 QY 361 nLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysG1 381
 Db 385 ARGCACTTTTACGTTCAATTTAACCATGCTCTAATCTGGGAAGGTGCTACTGTCGAGACCA 444
 QY 381 uArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLe 401
 Db 445 GAAGATATCTCTAC---GACGAATCGATCGACTCGCGGAATGCCAAAGAGCTCGCTGTCT 501
 QY 401 uValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAl 421
 Db 502 CTTCACTCAITTTGGTGGACGGTCTGTAAGGAGGGGTCCTTATCAGACAG-CCGTGGCG 560
 QY 421 aGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSe 441
 Db 561 AGCATATCGAAAAAGGGTTAGCCCTAGGCGACGGGATGTG-----CCTGCGTACAAGAA 614
 QY 441 x 441
 Db 615 C 615

RESULT 3
 AAS17845
 ID AAS17845 standard; cDNA; 3731 BP.
 XX AC AAS17845;
 XX DT 08-MAY-2002 (first entry)
 XX CC Tomato RNA-directed RNA polymerase (RDRP) cDNA from clone HF.
 XX KW Tomato; RDRP; RNA-directed RNA polymerase; in vitro transcription;
 KW cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
 KW transgenic plant; transgenic animal; cancer; viral infection;
 KW immunoprecipitation; immunolocalisation; RDRP/HF; mutant; gene therapy;
 XX SS.
 XX OS Lycopersicon esculentum.
 XX PH Key Location/Qualifiers
 FT mutation /tag= b
 FT mutation /tag= c
 FT CDS /tag= a
 FT /product= "RDRP protein"
 FT /EC_number= "2.7.7.48"
 FT mutation /tag= d
 FT mutation /tag= e
 FT mutation /tag= f
 FT mutation /tag= g
 XX US2001023067-A1.

```

PD 20-SEP-2001.
XX
XX
XX 08-FEB-2001; 2001US-00782874.
XX
XX 05-MAR-1997; 97US-00811583.
XX
XX (WASS/) WASSENEGGER M.
XX PA (RIED/) RIEDEL L.
XX
XX Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX WPI; 2001-595798/67.
XX
XX New nucleic acid molecule encoding a polypeptide having the enzymatic
XX PT activity of RNA-directed RNA polymerase, for modulating gene expression
XX PT and treating cancer and virus infection in human and animals.
XX
XX Example 3; Page; 34pp; English.
XX
XX This sequence represents a cDNA encoding the tomato RNA-directed RNA
XX CC polymerase (RDRP/HF) of the invention. The invention comprises the
XX CC nucleic acid and protein sequences of RDRP. The protein of the invention
XX CC can catalyze in vitro transcription of short single stranded RNAs into
XX CC DNA molecules, this transcription can be either primed by RNA or DNA
XX CC oligonucleotides or be unprimed. The protein may have cytosolic or
XX CC virucide activities. The sequences of the invention may be used in gene
XX CC therapy or as an RNA directed RNA synthesis inhibitor. The RDRP cDNA
XX CC sequence and a template nucleic acid molecule derived from a nucleic acid
XX CC molecule which causes a disease are useful for treating a disease caused
XX CC by the undesired expression or overexpression of a nucleic acid molecule
XX CC in a human, rat or mouse, by administering the molecules. This system can
XX CC be used in the preparation of a pharmaceutical composition and for
XX CC inhibiting expression of any desired gene by transferring the RDRP system
XX CC to organisms that either lack a comparable mechanism or do not
XX CC sufficiently express their own RDRP. An antibody or an antagonist or
XX CC inhibitor to the protein are useful for inhibiting RNA directed RNA
XX CC synthesis and for ensuring stable heterologous, gene expression in
XX CC transgenic organisms. The sequence is useful for probes and/or for the
XX CC control of gene expression. as primers for amplification of nucleic acid
XX CC molecules and as tools for the detection of expression of the cDNA
XX CC molecules. Additionally, nucleotide and protein sequences are useful for
XX CC suppression of undesired gene expression in humans and animals. The RDRP
XX CC is useful as a therapeutic agent for the control of cancer and virus
XX CC infection in humans and animals and the antibody is useful for
XX CC immunoprecipitation or immunolocalisation of the protein. Identification
XX CC of polypeptides interacting with it and screening expression libraries.
XX CC Note: This sequence is not shown in the specification but was created by
XX CC the indexer from the wild type Rdrp sequence shown in AAS17837 and the
XX CC information shown in table 1
XX
XX Sequence 3731 BP; 1062 A; 672 C; 948 G; 1149 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 5,028-15 Length: 3731
XX Score: 262.00 Matches: 127
XX Percent Similarity: 37.66% Conservative: 85
XX Best Local Similarity: 22.56% Mismatches: 222
XX Query Match: 8.70% Indels: 130
XX DB: 4 Gaps: 17
XX
XX
XX US-09-913-878a-2_COPY_710_1282 (1-573) x AAS17845 (1-3731)
XX
XX 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValIleuGlyLeuGly 20
XX Db 1772 GGAATGGTAAATATCTGGTCACTTGTCTCATAGAGTTGCCCAAAATGGCCCTTCAA 1831
XX
XX 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
XX Db 1832 TATACCCATCTGCTTCAGAGTTGTTATGTTGGATATAAGGTTGTGGGTGTTGAT 1891
XX
XX 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyProSerGlnArgIleAsp 60
XX Db 1892 CCGGATTCATCAATGAAG-----TTGCTTTTGAGAAAGAGCATGTCGAAATATGAA 1942

```

```

QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db 1943 TCAGACAAACATA-----AAGTTAGATGTCCTTGA--TGGAGCAATAT 1984
QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgIleAspLys 100
Db 1985 CAGCCTTGTTATCTTAATCGTCACTGATTAATCTCTCTCT----- 2026
QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db 2027 -----ACACTGGAGTGAAGATGAAGTTCGACAGACAGCAAAAGAGCTGTA 2077
QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrlGlySerTy 140
Db 2078 GATCAGCTTGATCTATCTTGCATGATCTTTTGAAGGCACAGAGGCTTTTGAATGATG 2137
QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db 2138 TCT-----CCTGGA 2146
QY 161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
Db 2147 GAGACACATAATATCTCAAGGCAATGCTAAACTGTGTATATAAGCCTGATGCTGAGCC 2206
QY 181 TyrlGluAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db 2207 TTTCTTTCATGATGTTGCCAACTTCCGCGCATCCAGTTGCTCGACTTGGACTAGA 2266
QY 201 LeuAsnIleArgValGlyArgSerAlaTyrlleTyMetIleAlaAspPheTrpGlyVal 220
Db 2267 TCAAGAATATTTATTCAAATGGAAGACAATGATGGGATGTTGGATGAATCCAGAAC 2326
QY 221 LeuGluGluAsnGluValHisValGlyPheSer-----SerLysPheArgAsp 236
Db 2327 TTGGAATATGCTCAGGTGTTGTTCACTTACTGTGCTGGACATGGAGAGTTTCTGAC 2386
QY 237 GluGluGluSerPhe-----ThrLeuLeuSerAspCysAsp 248
Db 2387 GATTACATCCATTATTAACAGCAGCATCCACCACAGTAATTTTCATTTCTGAAGGAAAT 2446
QY 249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValAlaVal 268
Db 2447 GTGGTGTGTCAAAAAATCCATGCTTGCATCTCGGTGATTCGTGTTTAAAGGCTGTA 2506
QY 269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
Db 2507 AATGTTGAGCGCTGCACCATGTTAGATTGTTGTGATTCCTCAGAAAGGAAAAAGA 2566
QY 289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrlAspGlyAspMetAlaTrpValCysTrp 308
Db 2567 CCTCATCCGAAATGAATGTTCTGGAGTGATTTGGATGGGGATATCTACTTTGTTCTGG 2626
QY 309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
Db 2627 GATCAGACATGATCCGCGCAAGGCAAGTCCAGCCGATGGAATATCTTCAGCAGCCAGC 2686
QY 329 ArgTyrlLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
Db 2687 ATACAGTTGGACCATGAT----- 2704
QY 349 SerAlaAlaLysGluGlnThrThrTyrlAspMetIleGlnLysSerPheHisPheAlaLeu 368
Db 2705 -----GTCACAAATGAGGAAGTTGAAGAGTACTTACCACACTATATT 2746
QY 369 GlnProAsnPheLeuGlyMetCysThrAsn-----TyrlLysGluArgLeuCys 384
Db 2747 GTGAATGACAGATTGGGAATCATAGCAAAATGCCATGTCGTATTTTGCAGACAGA----- 2800
QY 385 TyrlleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db 2801 ---GAACCTGATATGGCCATGAGTGTATCCATGCAAAAAAATCTGCTGAGCTCTTTCAATT 2857

```


Db 2138 TCT-----CCTGGA 2146
 Qy 161 SerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLys 180
 Db 2147 GAGAACACATAATCTCAAGCAATGCTAACTGGTTATAAGCTTGATGAGCC 2206
 Qy 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
 Db 2207 TTTCTTCAATGATGTGCAAACTTCGCGCATCCAAAGTCTCGATTTGGGACTAGA 2266
 Qy 201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
 Db 2267 TCAAGATATTTATTCATCAATGAAGAACATGATGGGATGTTTGGATCAATCCAGAAC 2326
 Qy 221 LeuGlnGluAsnGluValHisValGlyPheSer-----SerLysPheArgAsp 236
 Db 2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGTGTGGACATGGAGAGTTTCTGAC 2386
 Qy 237 GluGluGluSerPhe-----ThrLeuLeuSerAspCysAsp 248
 Db 2387 GATTATACATCCATTATTAACACAGATCCACCAAGATTAATTCATTCGAAGGAAAT 2446
 Qy 249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValAlaVal 268
 Db 2447 GTGGTTGTTCGCAAAAAATCCATCTTCGCATCCCTGGTGATATTCGTGTTTAAAGGCTGA 2506
 Qy 269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
 Db 2507 AATGTCGAGCGCTGCACACATGGTAGATTGTTGTTATTCCTCAGAAAGAAAGA 2566
 Qy 289 ProLeuAlaLysLysLeuSerGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
 Db 2567 CCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGGATATCTACTTTGTTGCTGG 2626
 Qy 309 AspProGluLeuValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
 Db 2627 GATCAGACATGATCCGCAAGCAAGTCCAGCCATGATATCTCCAGCACCCAGC 2686
 Qy 329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
 Db 2687 ATACAGTTGGACCATGAT----- 2704
 Qy 349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
 Db 2705 -----GTCACATGTGAAGAGTTGAAGAGTACTTCCACCACTATAT 2746
 Qy 369 GlnProAsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCys 384
 Db 2747 GTGAATGACAGTTTGGGAATCATAGCAATGCCATGCCATGCTGATTTGCAGACAGA----- 2800
 Qy 385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
 Db 2801 ---GAACCTGATATGCCATGATGATGATCCATGCAAAAACTTGTGAGCTCTTTTCAATT 2857
 Qy 405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
 Db 2858 GCAGTGGACTTCCAAAGACTGTGTTCGCGTGAAATACCATCT-----CAGTTGGC 2911
 Qy 425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
 Db 2912 -----CCTAAAGAAATPACCCAGACTTCATG 2935
 Qy 444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
 Db 2936 GATAAGCCGG-----ACAAGACC 2953
 Qy 464 AlaIleAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
 Db 2954 AGCTATATCTCAGAAAGATGATTGGAAAGCTTTTC-AGCAAGATGAAG----- 3001
 Qy 482 pThrGluAspGlyValAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePhe 502

Db 3002 -----GACAAAGCA-----CCTCAGGCTAGCTCTATCGGACCTTCACAAG 3042
 Qy 502 sGluIleSerAspLysSerArgSerSerAlaLeu----- 513
 Db 3043 AGATGTTGCAAGGAGATCATATGATCTGATGGAAGTTGATGGATTGAAGATTACAT 3102
 Qy 514 -----LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysG 528
 Db 3103 TGACGAGCTTTTGGACTACAAACTGAATATGACACAAAGCTGGTAAATTATGACTA 3162
 Qy 528 uTyrGly 530
 Db 3163 CTATGGC 3169
 RESULT 5
 AAD04370
 ID AAD04370 standard; cDNA; 3731 BP.
 XX
 AC AAD04370;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Tomato RNA-directed RNA polymerase (RdRP) cDNA.
 XX
 KW Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
 KW transgenic plant; tissue culture; plant breeding; therapy; C-protein; sp.
 XX
 OS Lycopersicon esculentum.
 XX
 FH Key Location/Qualifiers
 CDS 194..3538
 FT /*tag= a
 FT /product= "tomato C-protein having RNA-directed RNA
 FT polymerase (RdRP) activity"
 XX
 PN US6218142-B1.
 XX
 PD 17-APR-2001.
 XX
 PF 05-MAR-1997; 97US-00811583.
 XX
 PR 05-MAR-1997; 97US-00811583.
 XX
 PA (WASS/) WASSENEGGER M.
 PA (RIED/) RIEDEL L.
 XX
 PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
 XX
 DR WPI; 2001-289830/30.
 DR P-PSDB; AAE00897.
 XX
 PT New nucleic acid molecules encoding polypeptides with RNA-directed RNA
 PT polymerase enzymatic activity, useful in modulating gene expression in
 PT plants, humans and animals, as well as in plant cell/tissue cultures or
 PT plant breeding.
 XX
 PS Claim 1; Col 25-34; 31pp; English.
 XX
 CC The present sequence is a cDNA encoding tomato C-protein having RNA-
 CC directed RNA polymerase (RdRP) activity. This protein is capable of RNA-
 CC directed RNA synthesis, thus using RNA as a template for synthesizing
 CC complementary RNA molecules. RdRP nucleic acid is useful for modulating
 CC gene expression in plants, humans and animals. This may lead to various
 CC physiological, developmental and/or morphological changes. Transgenic
 CC plants containing RdRP nucleic acid is especially useful in plant cell or
 CC tissue cultures and in plant breeding. RdRP is useful in gene therapy,
 CC particularly for treating a disease that is caused by the undesirable
 CC expression or overexpression of a gene
 XX
 SQ Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.02e-15 Length: 3731

Score: 262.00 Matches: 127
Percent Similarity: 37.66% Conservative: 85
Best Local Similarity: 22.56% Mismatches: 222
Query Match: 8.70% Indels: 130
DB: 4 Gaps: 17

US-09-913-878A-2_COPY_710_1282 (1-573) x AAD04370 (1-3731)

| | | | |
|----|------|---|------|
| QY | 1 | GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly | 20 |
| DB | 1772 | GAATGGTAAATATCTCGTGACTTGTCTATAGAGTGGCTCAAAATGGCGCTTCAA | 1831 |
| QY | 21 | AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp | 40 |
| DB | 1832 | TATACCCATCTCGTTCAGATTCGTTATGCTGATATAAGGTGTTGGGTGTTGAT | 1891 |
| QY | 41 | ValAspAspThrGlyAspCysAspTrpIleGluThrTyProSerGlnArgIleTrpGlu | 60 |
| DB | 1892 | CCGGATTCATCAATGAAG-----TTGCTTTGAGAAGAGCATGTCGAATATGAA | 1942 |
| QY | 61 | CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu | 80 |
| DB | 1943 | TCAGACAACATA-----AAGTTAGATGTCCTTGA-----TGAGCAAAATAT | 1984 |
| QY | 81 | LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys | 100 |
| DB | 1985 | CAGCCTTGTTATCTTAATCGTCAACTGATACGCTCTTGTCT----- | 2026 |
| QY | 101 | VallLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer | 120 |
| DB | 2027 | -----ACATTGGAGTGAAGATGAAGTCTCGACAGCAAAAGAGCTGTA | 2077 |
| QY | 121 | GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyGlnSerTyr | 140 |
| DB | 2078 | GATCAGCTTGATCTATCTTGCTGATCTCTTTGAAGGCACAGAGGCTTGGAAATGATG | 2137 |
| QY | 141 | SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp | 160 |
| DB | 2138 | TCCT-----CCTGGA | 2146 |
| QY | 161 | SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLys | 180 |
| DB | 2147 | GAGAACACTAATATCTCAAGCAATGCTAAACTGTGGTATAAGCCTGATGCTGAGGCC | 2206 |
| QY | 181 | TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys | 200 |
| DB | 2207 | TTTCTTTCAATGATGTGCAAACTTCGCGCATCCAGTTCCTCGATTTGCGGACTAGA | 2266 |
| QY | 201 | LeuAsnIleArgValGlyArgSerAlaTyIleTyrMetIleAlaAspPheTrpGlyVal | 220 |
| DB | 2267 | TCAAGATATTTATTCCAATGGAAGAACATGATGGGATGTTTGGATCAATCCAGAAC | 2326 |
| QY | 221 | LeuGluGluAsnGluValHisValGlyPheSer-----SerLysPheArgAsp | 236 |
| DB | 2327 | TTGGAATATGCTCAGGTGTTTCTTCACTTACTGCTGCGACATGGAGAGTTCCTGAC | 2386 |
| QY | 237 | GluGluGluSerPhe-----ThrLeuLeuSerAspCysAsp | 248 |
| DB | 2387 | GATTTACATCCATTTAATACACAGATCCACCAACAGTAATTTCACTTCGAGGGAAT | 2446 |
| QY | 249 | ValLeuValAlaArgSerProLHisPheProSerAspIleGlnArgValArgAlaVal | 268 |
| DB | 2447 | GTGGTTGTTCGCAAAAATCCATGCTTCGATCCTCGGTGATATTCGTGTTTAAAGGCTGA | 2506 |
| QY | 269 | PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal | 288 |
| DB | 2507 | AATGTCGAGCGCTGCACACATGGTAGATTGTGTGTATTCCTCCAGAAAGAAAAGA | 2566 |
| QY | 289 | ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp | 308 |
| DB | 2567 | CCTCATCCGAATGAATGTTCTGGAGTGATTTGGATGGGATATCTACTTGTGTTGCTGG | 2626 |
| QY | 309 | AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer | 328 |

RESULT 6

| | | | |
|----------|----|---------------|--|
| AAH77704 | ID | AAH77704 | standard; cDNA; 3591 BP. |
| XX | XX | AAH77704; | |
| AC | AC | XX | |
| XX | XX | XX | |
| DT | DT | 13-NOV-2001 | (first entry) |
| XX | XX | XX | |
| DE | DE | XX | Nucleotide sequence of a plant SGS2 polypeptide. |
| KW | KW | XX | SGS2; RNA-dependent RNA polymerase; transgene silencing; |
| KW | KW | XX | transgene stability; crop plant; viral resistance; ss. |
| XX | XX | XX | Arabidopsis thaliana. |
| XX | XX | XX | |
| PH | PH | Key | Location/Qualifiers |
| FT | FT | CDS | 1..3591 |
| FT | FT | | /*tag= a |
| XX | XX | | /product= "SGS2" |
| PN | PN | FR2804128-A1. | |

| | | | |
|----|------|---|------|
| Db | 2627 | GATCAAGACATGATCCCGCCAGGCAAGTCCAGCCGATGGAATATCTCCAGCACCCAGC | 2686 |
| QY | 329 | ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly | 348 |
| Db | 2687 | ATAACAGTTGGACCATGAT | 2704 |
| QY | 349 | SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu | 368 |
| Db | 2705 | -----GTCACAATTGAGAGATTGAAGAGTACTTCACCAACTATATT | 2746 |
| QY | 369 | GlnProAsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCys | 384 |
| Db | 2747 | GTCAATGACAGATTGGGAATCATAGCAAAATGCCCATGCTGTATTTTCGACACAGA | 2800 |
| QY | 385 | TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn | 404 |
| Db | 2801 | ---GAACCTGATATGCCCATGAGTGCATGCAAAAACCTTGCTGAGCTCTTTCAATT | 2857 |
| QY | 405 | LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg | 424 |
| Db | 2858 | GCAGTGGACTTCCAAAAGACTGGTGTCCCGCTGAAATACCATCT-----CAGTTGCGC | 2911 |
| QY | 425 | ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr | 444 |
| Db | 2912 | -----CCTAAGAATACCCAGACTTCATG | 2935 |
| QY | 444 | pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr | 464 |
| Db | 2936 | GATAAGCCGG-----ACAAGACC | 2953 |
| QY | 464 | OAlaIleAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaLysAs | 482 |
| Db | 2954 | AGCTATATCTCAGAAAGAGATTATGAAAGCTTTTC-AGAAAGTGAAG----- | 3001 |
| QY | 482 | pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLys | 502 |
| Db | 3002 | -----GACAAAGCA-----CCTCAGGCTAGCTCTATCGCGACCTTCACAAG | 3042 |
| QY | 502 | sGluIleSerAspLysSerArgSerSerAlaLeu----- | 513 |
| Db | 3043 | AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT | 3102 |
| QY | 514 | -----LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysG1 | 528 |
| Db | 3103 | TGACGAAGCTTTTGACTACAAAACCTGAAATATGACACACAGCTGGTAATTTAATGGACTA | 3162 |
| QY | 528 | uTyrGly | 530 |
| Db | 3163 | CTATGGC | 3169 |


```

PD 27-JUL-2001.
XX
XX 26-JAN-2000; 2000FR-00001007.
XX
XX 26-JAN-2000; 2000FR-00001007.
XX
XX (RHOB-) RHOBIO.
XX (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX Beclin C, Elmayan T, Mourrain P, Vaucheret H;
PI
XX
XX WPI; 2001-543303/61.
DR
DR P-PSDB; AAG57235.
XX
XX
XX A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT polymerase and in transgene silencing, increases transgene stability and
PT expression in transgenic plants when it is inactivated.
XX
XX
XX Claim 4; Page 32-37; 46pp; French.
XX
XX The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-
CC dependent RNA polymerase and is involved in transgene silencing.
CC Inactivation of SGS2 is used to increase transgene stability and
CC expression in plants, particularly crop plants, especially maize, corn,
CC barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC Overexpression of SGS2 can be used to increase resistance to viral
CC infection in plants
XX
XX
XX Sequence 3591 BP; 1022 A; 714 C; 858 G; 997 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1.17e-14 Length: 3591
Score: 258.00 Matches: 140
Percent Similarity: 37.58% Conservative: 99
Best Local Similarity: 22.01% Mismatches: 247
Query Match: 8.57% Indels: 150
DB: 4 Gaps: 28
US-09-913-878A-2_COPY_710_1282 (1-573) x AAH77704 (1-3591)
Qy 1 GlyValGluArgMetSerArgSerValAlaLysArgAlaAspValLeuGlyLeuGly 20
Db 1759 GGAATTGGTACATCACACCTGACCTCGCTGACGAAGTAATGGAGAACCTTAAGTTG--- 1815
Qy 21 AspVal-----ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrp 37
Db 1816 GATGTGCATACACGCCCTTGCTGTTATCATCATACGTTACGCAGGTTTCAAAGGGGTT-- 1872
Qy 38 ValIleAspValaAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln-- 56
Db 1873 -----GTTCTCGTTGGCCATCAAAAGT 1896
Qy 57 -----ArgLysTrpGluCysAspPheValAspLysHisGlnArg 69
Db 1897 GATGGAATACGGGTAGCCCTTCGAGACAGATGAAAGAAGTCTCTTTTCCAACAT---ACG 1953
Qy 70 ThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 89
Db 1954 ATCTTGGAGATCTGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAATCGCAGATA 2010
Qy 90 LeuProValLeuGluAspArgAlaAspLysValLysMetArgGlnAlaIleGlyAsp 109
Db : : : : :
Qy 2011 ATTACCCTTCTATCC-----GTACAGGTGTCGCGAT 2043
Qy 110 ArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgPro 129
Db : : : : :
Qy 2044 GAAATATTCGGGATATGCAG-----GAATCCATGCTCTATAACTGAACCGCATC 2094
Qy 130 ValGluPheArgGlnTrpValTrpGluSerTyrSerSerArgAlaThrArgValSerHis 149
Db : : : : :
Qy 2095 CTTGATGATACAGATGTGGCATTTGAAATGTTCTCACGGCATCATGTGCT----- 2142
Qy 150 GlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeu 169

```

Db 3112 GCATGGGTCAAAATGTTTACACGGCGCAACTCATTTGGTCTCTTTGGCAATACAG 3171
 QY 479 AlaAlaLysAspThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
 Db 3172 GTGCAGAAAGAGGAGATTGTGACGGGTACATCTGGTCC-----ATGCCCAA 3222
 QY 498 TyrThrPhePheGlyGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThr 517
 Db 3223 TACACAGCAAGAAACAAGCGCAACTGAAGAAAGA----- 3258
 QY 518 LeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGlu 537
 Db 3259 CTGAGCATCTTATTAATCCCTTAAAGAGTTC---AGGAAGTATTTGAGGAACA 3315
 QY 538 MetArgAspSerLysAspProTyrProValArgValAsnGlnValTyrGluLys----- 555
 Db 3316 ATCCCTGACCATGAAATCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCT 3375
 QY 556 -----TrpCysAlaIleThr-----ProGluAlaMetAspLysSer 567
 Db 3376 TCAGCTTGTATCATGTAACTTACCATCCCGAGTGGGTGAAGAGTCT 3423

RESULT 7
 AAC63741
 ID AAC63741 standard; cDNA; 2816 BP.
 XX AAC63741;
 XX
 DT 07-FEB-2001 (first entry)
 XX
 DE Rice RNA-directed RNA polymerase P1S sequence, SEQ ID NO: 9.
 XX
 KW Rice; plant; RNA-directed RNA polymerase; P1S; gene mapping; gene marker;
 KW Plant virus resistance; plant breeding; ss.
 XX
 OS Oryza sativa.
 XX
 XN WO200060097-A1.
 XX
 PD 12-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US009105.
 XX
 PR 07-APR-1999; 99US-0128094P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J,
 PI Zhong G;
 XX
 DR WPI; 2000-679376/66.
 DR P-PSDB; AAB28532.
 XX
 PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
 PT for controlling gene expression and providing mechanisms to engineer
 PT plant virus resistance.
 XX
 PS Claim 2; Page 45-46; 62pp; English.
 XX
 CC The present sequence is one of a number of cDNA molecules which encode
 CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
 CC probes for genetically and physically mapping genes, and as markers for
 CC traits linked to those genes. They are useful for controlling gene
 CC expression and provide mechanisms to engineer plant virus resistance.
 CC They are also useful for plant breeding to develop lines with desired
 CC phenotypes
 XX
 SQ Sequence 2816 BP; 782 A; 572 C; 725 G; 736 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 1.66e-14

Length:

2816

Matches:

106

Score:

255.00

106

Percent Similarity: 39.33% Conservative: 82
 Best Local Similarity: 22.18% Mismatches: 188
 Query Match: 8.47% Indels: 102
 DB: 3 Gaps: 16

US-09-913-878a-2_COPY_710_1282 (1-573) x AAC63741 (1-2816)

QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgGlyLeuGlyLeuGly 20
 Db 673 GGAATTGGCAAAATTACCCAGATCTTGATTTGGAAGTTGCCGAGAGGCTGCACAG 732
 QY 21 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly---Met 36
 Db 733 GATAACCCGCCATCTCTTATCATGATCAGTTTGTGGCTTCAGGGTGTACAGTGTGTC 792
 QY 37 TrpValIleAspValAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln 56
 Db 793 TGG-----CAAGGACATGCTGATGGACACGGCTTTCTCGAGGCCAAGCATG 840
 QY 57 ArgGlyTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerVal 76
 Db 841 AGGAAGTTTGAGTCTAACCATTTG-----GTGTAGGGGTG 876
 QY 77 AlaSer-----GluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGlu 94
 Db 877 GTCTCTGGCAAAAGTTCCAGCCAGGATCTTTAAATCGACAGATTATAATATTGCTATCC 936
 QY 95 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsnAsp 114
 Db 937 TCA-----CTGAATGTCCTCCAGATTCTATCTTTTGGCAA 969
 QY 115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 134
 Db 970 ATCAAGACAGACCAGTCTTCTTAAC-----CTCAACAATATTCTATCAGACAGAT 1020
 QY 135 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 154
 Db 1021 GTTGTCTTTGAGGTTTTAAACAATTCATGTGCT----- 1053
 QY 155 LeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPhe 174
 Db 1054 -----GATGATGGAACACACTGCAGCATTTGATGCTAGTGTGCTTT 1095
 QY 175 AspProLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 194
 Db 1096 GAACCTAGAACTGACACACACTTGAAGCAATGCTCTTGGCTATAAGTGTGCACAATTG 1155
 QY 195 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 214
 Db 1156 CAGGATCTTTGGAAAAGCAGGATATTGTGCCAAGGAGGTGGTGTGATGGGCTGT 1215
 QY 215 AlaAspPheTrpGlyValLeuGluGlu----- 223
 Db 1216 CTTGTATGAGCTTGGGTTCTTTGAGCAAGGGCAGTGTCTTTATTCGGCAACAGTTCATCA 1275
 QY 224 ---AsnGluValHisValGlyPheSerSerLysPheArgAspGluGluSerPheThr 242
 Db 1276 TTGAATAGTATTATTGTTAAGCATGGGTCAAGATTTCATCAACAGATAAACAACAGAG 1335
 QY 243 LeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIle 262
 Db 1336 GTCATTTTGGGT---ACTGTGGTAATAGCAAGAATCCCTGTCTTCATCCAGGGGATGTC 1392
 QY 263 GlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePhe 282
 Db 1393 CGCATCTTGAACGAGTTGATGCGCCCAACTGCATCATCTGTTGATTGTTGGTGTTC 1452
 QY 283 SerThrLysGlyAspValProLeuAlaLysLeuSerGlyGlyAspTyrAspGlyAsp 302
 Db 1453 CCCAGAAAGTTCAGAGGCCACATGCTAACAGGCACTCTGGAGCGCATCTTGATGGGAT 1512
 QY 303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322
 Db 303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322

Db 1513 CTCCTACTTTGTGACATGGGATGAGAACTTATACCTCCAGGCAAGAGAGCTGGAACCCCT 1572
 Qy 323 LeuGlu-----ProAspLeuSerArgTyrLeuLysAspLysThrPheLys 339
 Db 1573 ATGGACTACTCCCACTGAAGCA-----AAA 1599
 Qy 340 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluInThrTyrAspMet 359
 Db 1600 CAACCTT-----CGCGCCCAAGTATCTCAACATGATATC 1632
 Qy 360 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 379
 Db 1633 ATGATGTTCTTTTAAAGAACATGATAGTGAATCTGTGTAGGATCTGTAATGCTCAT 1692
 Qy 380 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleLeuSer 399
 Db 1693 GTTGTTTCATGCTGATCTTAGCGAGTATGGTGCATGGATGAGAGTATTCACCTAGCT 1752
 Qy 400 SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer 419
 Db 1753 GAGCTAGCGAAGTCCCGTGGACTTCCCAAGACTGGC----- 1791
 Qy 420 TrpAlaGlnLeuArgGluLeuGlyGlyAlaLeuSer-----LeuProAsp 436
 Db 1792 -----AACTTCGATATATGCCACACACCTTTAAACCAAAAGTCTACCCCTGAC 1839
 Qy 437 ProMet-----TyrLysSerAspSerTrpLeuGlyArg 447
 Db 1840 TTCATGGGAAGAGATGGACAATCTTATAATCAGAAAGATCTTGGGAGG 1893

RESULT 8

AAC63742
 ID AAC63742 standard; cDNA; 3807 BP.

AC AAC63742;

DT 07-FEB-2001 (first entry)

XX Soybean RNA-directed RNA polymerase FIS contig, SEQ ID NO: 11.

DE Soybean; plant; RNA-directed RNA polymerase; FIS; gene mapping;
 KW gene marker; plant virus resistance; plant breeding; ss.

XX Glycine max.

XX W0200060097-A1.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US009105.

XX 07-APR-1999; 99US-0128094P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX (PION-) PIONEER HI-BRED INT INC.

XX Odell JT, Orozco EM, Wang Z, Sakai H, Cahoon RE, Wang J;
 XX Zhong G;

XX WPI; 2000-679376/66.

XX P-PSDB; AAB28533.

XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
 XX for controlling gene expression and providing mechanisms to engineer
 XX plant virus resistance.

XX Claim 2; Page 49-50; 62pp; English.

XX The present sequence is one of a number of cDNA molecules which encode
 XX plant RNA-directed RNA polymerase proteins. The sequences are useful as
 XX probes for genetically and physically mapping genes, and as markers for
 XX traits linked to those genes. They are useful for controlling gene
 XX expression and provide mechanisms to engineer plant virus resistance.

CC They are also useful for plant breeding to develop lines with desired
 CC phenotypes
 XX Sequence 3807 BP; 1132 A; 662 C; 836 G; 1185 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.: 6.85e-14 Length: 3807
 Score: 250.50 Matches: 132
 Percent Similarity: 38.25% Conservative: 91
 Best Local Similarity: 22.64% Mismatches: 240
 Query Match: 8.32% Indels: 121
 DB: 3 Gaps: 23

US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63742 (1-3807)

Qy 1 GlyValGlyArgMetSerArgSerValAlaLysArgAspValLeuGlyLeuGly 20
 Db 1818 GGAATGGGAAATAATCTCTTGAATTTGCCAGAAAGTGGTAAAAAATGCTGTATGAT 1877
 Qy 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
 Db 1878 TGCACCTCCATCTGCTTTCAGATTCGATATGGGTACAAAGAGTTGTGCGTGTGAC 1937
 Qy 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
 Db 1938 CCTAAATCATGCTACAAG-----TTATCACTCAGGAAGAGCATGCGGAAGTATGAT 1988
 Qy 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
 Db 1989 TCAGAT-----AACACAAAGTTAGATGTTTGGCC---CGTAGTAGTTT 2030
 Qy 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
 Db 2031 CAGCCATGTTATCTGAATCGGCAGTTAAATTTCTCTCTATCC----- 2072
 Qy 101 ValLysMetArgGlnAlaIleGlyAspArgGluLeuLeuAsnAspLeuGlnArgGlnPheSer 120
 Db 2073 -----ACTCTGTGTATCAAGGATGATGTTTTCAGAAAAAACAAGAGAACTGTT 2123
 Qy 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
 Db 2124 AATCACTGAACACTATPACTATACAGATTTCATTAAAGCTCAGGAAGTTCTGGACTTAATG 2183
 Qy 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
 Db 2184 TCTGCT-----GGAGAGATC----- 2198
 Qy 161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
 Db 2199 -----ACTAATGTTCTGAAGGAGATGCTCATTTGTGGATACAGCCTAATGAAGAACCA 2252
 Qy 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
 Db 2253 TTCCTTTCAATGATGCTTCAACCATTTAGGGCATCAAAACTTTTGGAAATTCGCACTTAA 2312
 Qy 201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
 Db 2313 TCTAGGATCTTTATTCAAAAGGAGCAATGATGGATGCTAGATGAACATAGAAC 2372
 Qy 221 LeuGluGluAsnGluValHisValGlyPheSerSer-----LysPheArgAsp 236
 Db 2373 CTAGAAATGATGTCAGATATTGTTTCACTTCTAACAATAGGCTCAGAGATCTATCTGAT 2432
 Qy 237 GluGluGluSerPheThrLeu-----LeuSerAspCysAspValLeuValAla 252
 Db 2433 GATTTTTTTTCATATGATTTTGCACAAAGAAATATATGTTAAAGTAAGTAGTGTGTAGCA 2492
 Qy 253 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 272
 Db 2493 AAAAAAYCCCTGCTTGCACCCAGGTGATGTCGTGTTTACAAAGCTGTGGATGTGCCAGAT 2552
 Qy 273 LeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAlaLys 292

Db 2553 TTGTACCACATGGTGGACTGTTGTTTCCCTCAAAAGGACCAAGACCTCATCCAAAT 2612
Qy LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluLeu 312
Db 2613 GAGTGTTCGGGAAGTATCTGATGGATGAGATATCTACTTTGTTGTTGGACCAATG 2672
Qy 313 ValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLys 332
Db 2673 ATT-----CCTTCTCGCCCAATGATCCA----- 2696
Qy 333 LysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys 352
Db 2697 ATGGACTATACCTGCTCCVCAACTCGGAATGGATCAT----- 2735
Qy 353 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 372
Db 2736 ---GATGTGATGATCAGAGAGTTGAGGAGTATTTTGCCAAATTACATAGTCAATGACAGT 2792
Qy 373 LeuGlyMetCysThrAsnTyr-----LysGluArgLeuCysTyrIle 386
Db 2793 CTGGGAATAATTGCCAATGCACACACTGCTTTTGCAGATAAAGACATTG----- 2843
Qy 387 AsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuVal 406
Db 2844 -----AAAGCAATGCTGATCAATGTGTAAAGCTTGCAAGTTGTTTCAACACGAGTT 2897
Qy 407 AspGlnSerLysGlnGlyLeuValPheAsnGluAlaSerTrpAlaGlnLeuArgGlu 426
Db 2898 GACTTTCCTAAACTGGTGTCCA-----GCTGTTATACCTCCTGAA 2939
Qy 427 Leu-----LeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSer 441
Db 2940 CTTTCATGTCAAGAATATCCTGACTTCATGGAGAGCCTGCAAAACCCACATACAAATCG 2999
Qy 442 AspSerTrpLeuGlyArg-----GlyGluProThrHis 452
Db 3000 CATACCTGATAGGAAGCTCTTTAGGAGTGAAGAATAATCAACAAGTCCGGCTCA 3059
Qy 453 IleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuAla 472
Db 3060 ATTACATCCTTCAAAAGTTG---GTTGCGAGAGACTCTTACGACCATGAATGGAAGTT 3116
Qy 473 -----PheHisAsnAlaMetLysAlaAlaLys 481
Db 3117 GATGGCTTCATGGATTATGTTGATGATGCTTTCTATCAAAACCAATATGACTACAAG 3176
Qy 482 AspThrGluAspGlyAlaHisPheTrpAspProAspLeuAla-----Ser 496
Db 3177 TT-GGAAATCTGATGGACTACTATGGATCAAACTCAAGCTGAATCCTCGGTGGAA 3235
Qy 497 TyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuPheThr 516
Db 3236 TATTATGAAAATGTCAAAATCTTTCAACAAAAGGAGGATCAGAAAGC----- 3283
Qy 517 ThrLeuLysAsnArgIleGly---GluValGluLysGluTyrGlyArg---LeuValLys 534
Db 3284 -----AATCATATGGCTGTGAGTCCCTTAGGAAAGAGCCAGGGCTTGTTCAA 3334
Qy 535 AsnLysGlu 537
Db 3335 TGAAGAACAG 3343
RESULT 9
ID AAC63739 standard; cDNA; 3737 BP.
XX
AC AAC63739;
XX
DT 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase EST sequence, SEQ ID NO: 5.
XX
KW Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;

gene mapping; gene marker; plant virus resistance; plant breeding; ss.
Zea mays.
WC200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US009105.
XX
PR 07-APR-1999; 99US-0128094P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
PI Zhong G;
XX
WPI; 2000-679376/66.
DR P-PSDB; AAB28530.
XX
PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance.
XX
PS Claim 2; Page 35-36; 62pp; English.
XX
CC The present sequence is one of a number of cDNA molecules which encode
CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC probes for genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes
XX
SQ Sequence 3737 BP; 1063 A; 760 C; 863 G; 1051 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 6,32e-13 Length: 3737
Score: 240.50 Matches: 137
Percent Similarity: 36.87% Conservative: 96
Best Local Similarity: 21.68% Mismatches: 258
Query Match: 7.99% Indels: 141
DB: 3 Gaps: 23
US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63739 (1-3737)
Qy 1 GlyValGlyArgMetSerArgSerValAla-----Lys 11
Db 1709 GGAGTTGGAAGATCTCAGCTAATTTTGCAGTGGAGTGGCTATGAAGTCAAAATTGAAA 1768
Qy 12 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 31
Db 1769 CGCTTT-----GCTCCTTCTGTTTTCAGATAAGGTATGCG 1804
Qy 32 SerAlaLysGlyMetTrpValIleAspValAspThrGlyAspGluAspTrpIleGlu 51
Db 1805 GGTTCACAGGTGTTCTCGCTGTA-----GATACAAGTCAATCATAGCTTTCT 1855
Qy 52 ThrTy-ProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 71
Db 1856 TTGAGAAAAGCATGTCAAAAGTTCAGTCCAGAAAATATC-----ACTCTT 1900
Qy 72 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuPro 91
Db 1901 GATGTCCTTGCATAC---AGCAAGTACCAACCATGCTTCTGATCGGCGATTGATTACT 1957
Qy 92 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 111
Db 1958 CTTCTCTCA-----ACACTTGGGCTTAGCATATGTC 1990
Qy 112 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 131

Db 1991 TTTGAGCTAAGCAGAGGAGGCTTAAGGCAG-----TTGACAGCAATGGTAACT 2041
 Qy 132 PheArgGlnTyrValTyrGlnSerTyrSerSerArgAlaThrArgValSerHisGlyArg 151
 Db 2042 -----GNACACAGGCTGCTCGTGAAGCAGTTGAACCTT 2074
 Qy 152 ValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsn 171
 Db 2075 ATGCCCATG-----GGACAGGTAACCAATGTAGTTAAAGAAATTTGTGTCA 2119
 Qy 172 SerGlyPheAspProLysGlnLysTyrLeuGlnAspIleAlaTyrAspLeuGlnLys 191
 Db 2120 TGTGGTACAGCCTGATCATGAGCCATATCTTCCATGCTGCTCAAACTTTTAGACA 2179
 Qy 192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 211
 Db 2180 TCCAAAGCTTCTAGAGTTGAACCAAGTCAAGGATATTATCATCACAGGGGGCAGCAATG 2239
 Qy 212 TyrMetIleAlaAspPheTyrGlyValLeuGluGlnAsnGluValHisValGlyPheSer 231
 Db 2240 ATGGGTTGCTGATGAACCTGCACACTTAAGTACGGCCAGGATATTCTGTCACAGCTTCT 2299
 Qy 232 SerLysPheArgAspGluGlnGluSerPheThrLeuLeuSerAspCysAspValLeuVal 251
 Db 2300 TACAGTGCAGATGACCATCGCAAG-----GTCTGTGTAACTGGAAGAGTAGTTGTC 2350
 Qy 252 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 271
 Db 2351 GCCAAATTCCTGTCTCCACCTGGTGACATACCGGGTCTCCAGCTGTGTGATGTTCT 2410
 Qy 272 GluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAla 291
 Db 2411 GCTCTGCACCATGTTTCACTGTGTGCTTTCCACAGCAGCGGACCAAGCGCCACCT 2470
 Qy 292 LysLeuLeuSerGlyAspTyrAspGlyAspMetAlaTyrValCysTrpAspProGlu 311
 Db 2471 AATGAGTGTTCAGGGAGTACTTGTATGGGACATATATTTGTTCTTGGGATCCACAT 2530
 Qy 312 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLysSerArgTyrLeu 331
 Db 2531 CTATTCCAAAGTCGTTGGTGATCCTATGGACTATATCCAGCTTCAGCAGAAACATTA 2590
 Qy 332 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 351
 Db 2591 GACCATGAT----- 2599
 Qy 352 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371
 Db 2600 -----GTCACTATTGAGGAGTACAGGAGTACTTCAAACTACATAGTTAATGAG 2650
 Qy 372 PheLeuGlyMetCysThrAsnTyr-----LysGluArgLeuCysTyr 385
 Db 2651 AGCTTTGGGATTATCGCCCAATGGCATGCTGCTTTTACAGATCAGGAACGTATG----- 2704
 Qy 386 IleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeu 405
 Db 2705 -----AAAGCTGAGATCCACCGTGGTTCACCTGCTCAAGCTCTTCTCTATAGCT 2755
 Qy 406 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 425
 Db 2756 GTCGATTCCCAAAGACTGGAGTGGCG-----GCTCTGATTCCACAT 2797
 Qy 426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet----- 438
 Db 2798 GAGCTA-----CATGTCAAGGAGTATCTGACTTTCATGGAGAACTCGACAAAGTCACC 2851
 Qy 439 TyrLysSerAspSerTrpLeuGlyArg-----GlyGluProThrHisIle 453
 Db 2852 TATGAATCAAGGGTGTGATCGGGAAGCTCTATAGGAAATAAAGAAAGCACACACACAC 2911
 Qy 454 IleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu-----Glu 471
 Db 2912 ATAAAGCACTTCACGAGGGAAGTGGCAAGCGGCTTTATGACCCGATTGTGATTGAT 2971

Qy 472 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 491
 Db 2972 GGCTATCAAGATTACATTACTAGAGCTATAGAGTTTCAAGGAAGAGTACGATTTCAGGCTG 3031
 Qy 492 ProAspLeuAlaSerTyrTyrThrPhePheLysGlu-----IleSerAspLysSer 508
 Db 3032 GGTAATCTTATGACCACTATGGCATAAAGTGAAGCTGAGATAATAAGTGGATGATTT 3091
 Qy 509 ArgSerSerAlaLeuLeuPheThr-----ThrLeuLysAsnArgIle 522
 Db 3092 CTAAGATGCAAGAAATTTCAACAGAGTAGTGTGATGCAATTAAGATGGCGGTG 3151
 Qy 523 GlyGluValGluLysGlu-----TyrGlyArgLeuValLysAsnLysGluMet 538
 Db 3152 AGATCTTTTGGAGAAAGAGCTAGGTGCTGCTTCAATGAGATGAGCAGGAGGATGGC 3211
 Qy 539 ArgAspSerLysAspPro-----TyrProValArgValAsnGlnValTyr 553
 Db 3212 CAAGTGCCTGAGGAGGAGGCTCTGCTGTGTGATCATCTTACTTATCATCAGCAGTAC 3271
 Qy 554 -----GluLysTrpCysAla 558
 Db 3272 TGGGCGCAGCTACAATGAAGGTTATGATCGGCCCATCTTATTAGCTTCCCATGGTGGTA 3331
 Qy 559 IleThrProGluAlaMetAspLysSerGlyAlaAsn 570
 Db 3332 TATGACAGCTTGTGGCCATCAAGCAGGAGGAGAT 3367
 RESULT 10
 AAC63740
 ID AAC63740 standard; cDNA; 3901 BP.
 XX AAC63740;
 AC AAC63740;
 XX 07-FEB-2001 (first entry)
 DT 07-FEB-2001 (first entry)
 XX Maize RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 7.
 DE Maize; plant; RNA-directed RNA polymerase; FIS; gene mapping;
 KW gene marker; plant virus resistance; plant breeding; ss.
 KW Zea mays.
 OS
 FN WO200060097-Al.
 XX 12-OCT-2000.
 PD 06-APR-2000; 2000WO-US009105.
 XX 07-APR-1999; 99US-0128094P.
 PR (DUPO) DU FONT DE NEMOURS & CO E I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
 PI Zhong G;
 XX WPI; 2000-679376/66.
 DR P-ESDB; AAB28531.
 XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
 PT for controlling gene expression and providing mechanisms to engineer
 PT plant virus resistance.
 XX Claim 2; Page 40-41; 62pp; English.
 PS The present sequence is one of a number of cDNA molecules which encode
 CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
 CC probes for genetically and physically mapping genes, and as markers for
 CC traits linked to those genes. They are useful for controlling gene
 CC expression and provide mechanisms to engineer plant virus resistance.
 CC They are also useful for plant breeding to develop lines with desired

CC phenotypes
XX Sequence 3901 BP; 928 A; 975 C; 1058 G; 935 T; 0 U; 5 Other;
SQ Alignment Scores:
Pred. No.: 1,058-12 Length: 3901
Score: 238.50 Matches: 138
Percent Similarity: 36.48% Conservative: 98
Best Local Similarity: 21.33% Mismatches: 248
Query Match: 7.92% Indels: 163
DB: 3 Gaps: 26

US-09-913-878a-2_COPY_710_1282 (1-573) x AAC63740 (1-3901)

QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
DB 1857 GGAATTGGCAAGATTACCTGGGACCTTGCACCTCGAAGTTGCTCAGAAGCTGCAATTGACA 1916

QY 21 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIle 39
DB 1917 GATAATCCCATCTGCTTACAGATTAGGTATGAGGCTTCAAGGCT---GTTATA 1970

QY 40 AspVal---AspAspThrGlyAspGluAspTrpIleGluThrTrpProSerGlnArgLys 58
DB 1971 TCTGCTGGGAAGGAAAAATGATGGGATACGACTTTCCCTGAGCGCGAGCATGCCAAG 2030

QY 59 TrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSer 78
DB 2031 TTTGAG---TCTAACCATACTGTGTTAGAGGTGGTCTCG---TGGACA 2072

QY 79 GluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArg 98
DB 2073 AAGTTTCAGCCAGGAGTCTTAATCGTCAGATTATTACATTACTGTCC--- 2120

QY 99 AspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGln 118
DB 2120 ----- 2120

QY 119 PheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTrpGlu 138
DB 2121 -----TCTTGNATGTCGCGATGCTATCTTTGCTCAAGTCAGGAA 2162

QY 139 SerTyrSerArgAlaThrArgVal---SerHisGlyArgValProPhe-----Leu 155
DB 2163 GCCATGTTATCTAATCTCAACAATATTTGTCAGACTCTGATGTTGCTTTTGACATTGTA 2222

QY 156 AlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAsp 175
DB 2223 ACCGCTCTTGTGCTGACCAAGAACCTCCAGCAGCTAGTGTTGAGTGTGCTGGCATTTCA 2282

QY 176 ProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAsp 195
DB 2283 CTTGAACTGAGCCACACCTGAAGCAATGCTGTAGTATAAGGTCCTCAGCTGCTA 2342

QY 196 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAla 215
DB 2343 GTCCTTTTGGAGAACCAAGGATTTTGTGCCAAGGAAGGTGGTGTGATGGGCTGCGCTT 2402

QY 216 AspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArg 235
DB 2403 GATGAATGGGATGCTTGGACCAAGGCGAGTCTTATCCGGGCATCATCT----- 2453

QY 236 AspGluGluGluSerPheThrLeuLeuSerAspCys----- 247
DB 2454 -----CCATCACTCAATAATTTGCTGTGTAAGTATGGATCAAGATTG 2495

QY 248 -----AspValLeuValAlaArgSerPro 255
DB 2496 TCTGAGCAACAAACAAATGCAGAGCACCTTCTGGGTACTATCGTAATGGCAAGATCCA 2555

QY 256 AlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSer 275
DB 2556 TGCCTTCATCCAGGGGATGTCGAATCCTTGAAGCTGTGTGATGTGCTGAAGTGCATCAC 2615

QY 276 LeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSer 295
DB 2616 CTTGTTGATGCTGCTTCCCAAGAAAGGTGAGAGGCCGACGCAATGAGCATCT 2675

QY 296 GlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGly 315
DB 2676 GGGAGTGATCTTGTATGGGATCTATCTTCTGTAACATGGGATGAAACCTTATACACCT 2735

QY 316 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysLysLys 335
DB 2736 GGTAAAAAGAGTTGGAAACCTATGGAC-----TACTCCCCAGCTGAAGCA 2780

QY 336 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 355
DB 2781 AAACAACTCCACGGCAGTAGTATCCCAACAT----- 2810

QY 356 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375
DB 2811 -----GATATGTT-----GGTTTC-----TTCITGGAAGAC 2837

QY 376 CysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSer----- 389
DB 2838 ATGGTAAAT-----GAGAAACTGGGTCCATAAAGCAATGCTCATGTTGTTCAACGTGAT 2891

QY 390 -----ValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
DB 2892 ATGAGCGATATGGAGCAATGGATGAGAGTGTATTCAGTTGGCAGAACTAGCAGCACT 2951

QY 405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
DB 2952 GCTGTGGACTTCCCAAGACAGCAAAATTGTGTCATGCCA-----GCATCCCTTCGA 3005

QY 425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet---TyrLysSerAspSer 443
DB 3006 CCAAAATATATCTCTGACTTTCATGGAAAGGAGTATCTCTCTATAGATCAGAGAAG 3065

QY 444 TrpLeuGlyArg----- 447
DB 3066 ATCCTTGAAGGCTTTATCGGTCAATCCAAGAGCCTCCAGCGATGATTGTTGTTCCAGAA 3125

QY 448 -----GlyGluProThrHisIleIle 454
DB 3126 GAACTTGCACATCAACAATCTGCTTATGATGCAGATATGGAAGTTGCTGCTGCAGCT 3185

QY 455 AspTyrLeuLysPheSer---IleAlaArgProAlaIleAspLysGluLeuGluAlaPhe 473
DB 3186 GATTTTCTCTCAGTGTCTGGCAGTGCAGTCTCATATGAACACAACTGACCGACTG 3245

QY 474 HisAsnAlaMetLysAlaAlaLysAspThrGlu---AspGlyAlaHisPheTrpAspPro 492
DB 3246 CTCACCAATATGGCGTGGCAGTGAAGCAGAGCTTGTGACAGACATATATGG----- 3299

QY 493 AspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAla 512
DB 3300 TGCTTCCCAAGTAC-----AGCAGCAGAGGAGGAGGGG 3332

QY 513 LeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeu 532
DB 3333 GACATAAAGGAGAGGTTGAAGAAATGCATACTATGCTCTTTCACAGGAGTTT---AGGAGC 3389

QY 533 ValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGlnVal 552
DB 3390 ATTTTCGAAGCATTTGTGACAGATCAACAGAGATCTCTGTATGTAGTGAAGAAATCGGTTT 3449

QY 553 TyrGlu-----LysTrpCysAlaIleThr-----ProGluAlaMetAspLys 566
DB 3450 TACGAGATGAAGGCTCCGCTTGGTACCAAGTAACCTACCATCTCTGTAATGGGTCCAGAG 3509

QY 567 SerGlyAlaAsnTyrAspSer 573
DB 3510 TCAAGGGAATGTTCAAGTCT 3530


```

PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 32824; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at fcp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 4818 BP; 1368 A; 1145 C; 1214 G; 1091 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0.000852 Length: 4818
XX Score: 148.50 Matches: 137
XX Percent Similarity: 34.38% Conservatives: 93
XX Best Local Similarity: 20.48% Mismatches: 250
XX Query Match: 4.93% Indels: 189
XX DB: 4 Gaps: 33
XX
XX US-09-913-878A-2_COPY_710_1282 (1-573) x ABL27117 (1-4818)
XX
XX QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGly----- 18
XX
XX DB 25 GGCAGCCGAAGGACTCGAAGATCCCTAGCTTCGGGTGGTGGTATCTATTGGCCCTPAT 84
XX
XX QY 19 -----LeuGlyAspValProSerAlaValGlnGlyArgPhe----- 30
XX
XX DB 85 ATCTATTGCATCACTTCTATGATAATGATTGATGCGGTGCGGGATCAGTTTGTGAAG 144
XX
XX QY 31 -----GlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAsp 46
XX
XX DB 145 CGGCTTGAAGAGTGGGTACGAGGTGCCATGTTCTGTCATCGACTTGGATAGGAAGTTGCG 204
XX
XX QY 47 GluAspTrpIleGlnThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLys 66
XX
XX DB 205 GAG-----GCCAATTATCAGTTCACCAAGAG----- 231
XX
XX QY 67 HisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGly---Leu 85
XX
XX DB 232 ---TCGAGGACCAAGAAAGCTGGATCTCTG-----GACATGAAGTCTCTTAGATCCATTG 282
XX
XX QY 86 AsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValIleMetArgGln 105
XX
XX DB 283 AACATT---GTCTCGCAATGCAATTCAGAACTGTGTGGAAGAGTTTGACAGAAATACGAAG 339
XX
XX QY 106 AlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAla 125
XX
XX DB 340 CAGATTGAGTCAGAGCGCAATTTCGACCTGTACTCTCTACTGCGCAACGA----- 387
XX
XX 126 LeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerArgAlaThr 145

```

| | | | |
|--|-----|--|------|
| 1363 | Ddb | GTCTACCACCGCTGTCTGGAGTATAATCTTTTGGCAGCAGTACTTTTACATCCGGGTATATC | 1421 |
| 433 | Qy | SerLeuProAspProMetTyrLysSerAspSerTyrLeuGlyArgGlyGluProThrHis | 452 |
| 1423 | Ddb | -----AACAGCTCACTGAAGTGTCTCAGCCCTACAAATGGTGTGATGTGGAACCGATCTAT | 1473 |
| 452 | Qy | ----- | 452 |
| 1474 | Ddb | CTGCCCAAGTTTCCGAACCTCTATCTCACGGATGACTCGGTTATGCAACGAATTATCGAG | 1533 |
| 453 | Qy | ---IleIleAspTyrIleuLysPheSerIleAlaAaGProAlaIleAspLysGluLeuGlu | 471 |
| 1534 | Ddb | CTGCTCAGGACTACGATGACTTCAGCCTGGAGGAGTGCAGCCCAATGTCCATCTCTAT | 1593 |
| 472 | Qy | AlaPheHisAsnAlaMetLysAlaLaLysAspThrGluAspGlyAla----- | 487 |
| 1594 | Ddb | ACTTTTCCGAAGTCACTAAACGAGGATTTCGAACACAGGAAATCAACACAGTCATCCCAACA | 1653 |
| 488 | Qy | -----HisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSer | 505 |
| 1654 | Ddb | CGGCTCTGCTTCAGGAAATTTCACTGTACGAAATGGAGAATTTTCTAAGAGACTTG--- | 1710 |
| 506 | Qy | AspLysSerArgSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluVal | 525 |
| 1711 | Ddb | -----GTCACCCCAAAAGGTTCAACGAGGTG | 1737 |
| 526 | Qy | GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp-----SerLys | 542 |
| 1738 | Ddb | ACCGAAAGTTAT-----TTGTCAACGACAAACGAGAGCGCTGATAATATTGCGACCCAA | 1791 |
| 543 | Qy | AspProTyrProValArgValAsnGln | 551 |
| 1792 | Ddb | GATCAATTCACTAGTGGTGCACGAG | 1818 |
| RESULT 14 | | | |
| ABU71773 | | | |
| ID ABL71773 standard; cDNA; 282 BP. | | | |
| XX | | | |
| AC ABL71773; | | | |
| XX | | | |
| DT 14-VAY-2002 (first entry) | | | |
| XX | | | |
| DE Corn tassell-derived polynucleotide (cdps) SEQ ID NO:1147. | | | |
| XX | | | |
| KW Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPS; | | | |
| KW inheritance; characteristic; growth; development; disease resistance; | | | |
| KW environmental adaptability; quality; yield; molecular marker; | | | |
| KW multigene trait; plant breeding; corn tassell; gene; ss. | | | |
| XX | | | |
| OS Zea mays. | | | |
| XX | | | |
| PN US2001051335-A1. | | | |
| XX | | | |
| PD 13-DEC-2001. | | | |
| XX | | | |
| PF 16-APR-1999; 99US-00294093. | | | |
| XX | | | |
| PP 21-APR-1998; 98US-0082567P. | | | |
| XX | | | |
| PA (LALG/) LALGUDI R V. | | | |
| PA (ITOL/) ITO L Y. | | | |
| PA (SHER/) SHERMAN B K. | | | |
| XX | | | |
| PI Lalgudi RV, Ito LY, Sherman BK; | | | |
| XX | | | |
| WPI; 2002-163647/21. | | | |
| XX | | | |
| PT Novel purified corn tassell-derived polynucleotide useful for determining | | | |
| PT altered gene expression, to recover regulatory elements and to follow | | | |
| PT inheritance of desirable characteristics through hybrid breeding | | | |
| PT programs. | | | |
| XX | | | |
| PS Claim 1; SEQ ID NO 1147; 201bp; English. | | | |

```

XX CC The present sequence describes a purified corn tassel-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76832. The cdps sequences
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (I) are also useful in
CC the evaluation, and alteration of desired characteristics associated with
CC growth and development, disease resistance, environmental adaptability,
CC quality and yield, and as molecular markers for studying inheritance of
CC multigenic traits in a plant breeding program. (I) can be used to produce
CC a tassel-specific profile of gene transcription, a transcript image, to
CC clone regulatory elements for use in transformation vectors, to express a
CC polypeptide, to identify, isolate or extend identical or related corn
CC tassel nucleic acid sequences from DNA libraries, in nucleic acid
CC hybridisation or amplification technologies, as query sequences to
CC determine homology of known sequences, as probe for use in Southern or
CC Northern hybridisation, and to identify the presence of and/or to
CC determine the degree of similarity between two (or more) nucleic acid
CC sequences
XX SQ Sequence 282 BP; 78 A; 67 C; 70 G; 67 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2.04e-05      Length:      282
Score:          148.00        Matches:    26
Percent Similarity: 64.62%     Conservative: 16
Best Local Similarity: 40.00%   Mismatches: 23
Query Match:     4.92%       Indels:      0
DB:              6           Gaps:        0

US-09-913-878A-2_COPY_710_1282 (1-573) x ABL71773 (1-282)

Qy      249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
         : ::::: ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      21 ATCGTAAATGGCAAGAACCTCATGCCCTTCATCCAGGGATGTCCGAATCCTTGAACTGTT 80
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Qy      269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      81 GATGTGCCTGAACCTCATCACCTTGTGATTGCTTGTCTTCCCAGAAGAGTGAGAGG 140
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Qy      289 ProLeuAlaLysLeuLysSerGlyGlyAspTyRaspGlyAspMetAlaTrpValCysTrp 308
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      141 CGCAGCGCAAGTAAGCATCTGGGAGTAGTCTTGATGGGATCTATATCTTCGTAACATGG 200
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Qy      309 AppProGluIleVal 313
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db      201 GATGAAGACCTTATA 215
         : ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

RESULT 15
ABL71722
ID      ABL71722 standard; cDNA; 274 BP.
XX AC      ABL71722;
XX DT      14-MAY-2002 (first entry)
XX DE      Corn tassel-derived polynucleotide (cdps) SEQ ID NO:1096.
XX KW      Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
KW      inheritance; characteristic; growth; development; disease resistance;
KW      environmental adaptability; quality; yield; molecular marker;
KW      multigenic trait; plant breeding; corn tassel; gene; ss.
XX OS      Zea mays.
XX FN      US2001051335-A1.
XX PD      13-DEC-2001.
XX PF      16-APR-1999; 99US-00294093.
XX XX

```

```

PR 21-APR-1998; 98US-0082567P.
XX
XX (LALG/) LALGUDI R V.
PA (ITOL/) ITO L Y.
PA (SHER/) SHERMAN B K.
XX
XX Lalgudi RV, Ito LY, Sherman BK;
PI
XX WPI; 2002-163647/21.
DR
XX
XX Novel purified corn tassell-derived polynucleotide useful for determining
PT altered gene expression, to recover regulatory elements and to follow
PT inheritance of desirable characteristics through hybrid breeding
PT programs.
XX
XX Claim 1; SEQ ID NO 1096; 201pp; English.
PS
XX The present sequence describes a purified corn tassell-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cdps sequences
CC encode corn tassell-derived polypeptides (CDPs). The cdps sequences (I)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (I) are also useful in
CC the evaluation, and alteration of desired characteristics associated with
CC growth and development, disease resistance, environmental adaptability,
CC quality and yield, and as molecular markers for studying inheritance of
CC multi-gene traits in a plant breeding program. (I) can be used to produce
CC a tassell-specific profile of gene transcription, a transcript image, to
CC clone regulatory elements for use in transformation vectors, to express a
CC polypeptide, to identify, isolate or extend identical or related corn
CC tassell nucleic acid sequences from DNA libraries, in nucleic acid
CC hybridisation or amplification technologies, as query sequences to
CC determine homology of known sequences, as probe for use in Southern or
CC Northern hybridisation, and to identify the presence of and/or to
CC determine the degree of similarity between two (or more) nucleic acid
CC sequences
XX
SQ Sequence 274 BP; 68 A; 58 C; 72 G; 76 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.000715 Length: 274
Score: 132.00 Matches: 25
Percent Similarity: 58.67% Conservative: 19
Best Local Similarity: 33.33% Mismatches: 29
Query Match: 4.39% Indels: 2
DB: Gaps: 1

US-09-913-878A-2_COPY_710_1282 (1-573) x ABL71722 (1-274)
QY 249 ValLeuValAlaAAGSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
Db |||.....|||.....|||.....|||.....|||.....|||.....|||
2 GTTCAATCACAAACCCCTTGTCTCCATCCTGGTGATGTCAGACTTGNAGCTGTA 51
QY 269 PheLysProGluLeu-----HisSerLeuLysAspValIleIlePheSerThrLysGly 286
Db |||.....|||.....|||.....|||.....|||.....|||.....|||
62 TATGACCCCTGGATGGATGCTAGGGGTCCTTATTGATTGTGTTCTTCACAGAGGG 121
QY 287 AspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpVal 306
Db |||.....|||.....|||.....|||.....|||.....|||.....|||
122 GAAGGCGCTCATCCGAATGATGCTCCGGGGGGGATTGGATGGACCTTCCTTTATT 181
QY 307 CysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 321
Db |||.....|||.....|||.....|||.....|||.....|||.....|||
182 ACTTGGGATGACAAACTGATCCGGAGAGAGGTTGATGCACCTATG 226

```

Search completed: March 31, 2004, 07:26:40
Job time : 385.609 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 06:31:56 ; Search time 65.8587 Seconds

(without alignments)

4828.317 Million cell updates/sec

Title: US-09-913-878a-2_COPY_710_1282

Perfect score: 3010

Sequence: 1 GVGRMSRVAKIRIVLGLG.....EKMCATPEAMDKGANYDS 573

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082009_8395/app.query.fasta_1.2254
-DB=Issued Patents NA -QWMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=prc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878 -CGCN_1_82@runat_29032004_082009_8395 -NCPU=6 -ICPU=3
-NO WMAP -LARGSEQ=0 -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6C COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 262 | 8.7 | 3731 | 3 | US-08-811-583-1 |
| 2 | 124 | 4.1 | 274 | 4 | US-09-313-294A-5730 |
| 3 | 117.5 | 3.9 | 13977 | 4 | US-09-484-970B-60 |
| 4 | 114 | 3.8 | 1491 | 4 | US-09-543-681A-2293 |
| 5 | 113 | 3.8 | 8494 | 4 | US-08-961-527-163 |
| 6 | 112.5 | 3.7 | 6702 | 3 | US-08-987-439-2 |
| 7 | 111 | 3.7 | 12311 | 3 | US-08-750-717-1 |
| 8 | 110.5 | 3.7 | 6909 | 2 | US-08-804-196-1 |
| 9 | 110.5 | 3.7 | 6909 | 2 | US-08-658-340-1 |
| 10 | 110.5 | 3.7 | 6909 | 3 | US-08-746-111-26 |
| 11 | 110.5 | 3.7 | 6909 | 4 | US-08-454-353A-1 |
| 12 | 110.5 | 3.7 | 6909 | 4 | US-09-165-019-1 |

| | | | | | | |
|----|-------|-----|-------|---|---------------------|---------------------|
| 13 | 107.5 | 3.6 | 2077 | 4 | US-09-620-312D-985 | Sequence 985, App |
| 14 | 107.5 | 3.6 | 5307 | 4 | US-09-489-039A-4533 | Sequence 4533, Ap |
| 15 | 105.5 | 3.5 | 15462 | 3 | US-09-073-492-1 | Sequence 1, Appli |
| 16 | 104 | 3.5 | 3648 | 2 | US-09-134-001C-2482 | Sequence 2482, Ap |
| 17 | 103.5 | 3.4 | 6360 | 2 | US-08-841-349-3 | Sequence 3, Appli |
| 18 | 103.5 | 3.4 | 6960 | 4 | US-09-431-184A-3 | Sequence 3, Appli |
| 19 | 103.5 | 3.4 | 8176 | 2 | US-08-841-349-5 | Sequence 5, Appli |
| 20 | 103.5 | 3.4 | 8176 | 4 | US-09-431-184A-5 | Sequence 5, Appli |
| 21 | 103 | 3.4 | 1724 | 4 | US-09-117-860-36 | Sequence 36, Appli |
| 22 | 103 | 3.4 | 4495 | 4 | US-09-620-312D-352 | Sequence 352, Appli |
| 23 | 103 | 3.4 | 10320 | 4 | US-09-091-501B-9 | Sequence 9, Appli |
| 24 | 102 | 3.4 | 1263 | 4 | US-09-252-991A-2381 | Sequence 2381, Ap |
| 25 | 102 | 3.4 | 2814 | 4 | US-09-252-991A-2537 | Sequence 2537, Ap |
| 26 | 101.5 | 3.4 | 7881 | 2 | US-08-751-189-1 | Sequence 1, Appli |
| 27 | 101.5 | 3.4 | 7881 | 2 | US-09-060-836-1 | Sequence 1, Appli |
| 28 | 101.5 | 3.4 | 7881 | 3 | US-09-184-445-1 | Sequence 1, Appli |
| 29 | 101 | 3.4 | 12284 | 2 | US-08-876-991-1 | Sequence 1, Appli |
| 30 | 101 | 3.4 | 12284 | 2 | US-09-059-853-1 | Sequence 1, Appli |
| 31 | 100.5 | 3.3 | 8351 | 1 | US-08-198-445B-14 | Sequence 14, Appli |
| 32 | 100.5 | 3.3 | 8351 | 2 | US-08-870-693-14 | Sequence 14, Appli |
| 33 | 100.5 | 3.3 | 19307 | 3 | US-08-836-022A-10 | Sequence 10, Appli |
| 34 | 100.5 | 3.3 | 19307 | 3 | US-09-427-048A-10 | Sequence 10, Appli |
| 35 | 100 | 3.3 | 7453 | 4 | US-09-620-312D-248 | Sequence 248, App |
| 36 | 100 | 3.3 | 7501 | 4 | US-09-620-312D-249 | Sequence 249, App |
| 37 | 100 | 3.3 | 7596 | 4 | US-09-023-655-1463 | Sequence 1463, Ap |
| 38 | 100 | 3.3 | 11283 | 2 | US-08-603-753D-3 | Sequence 3, Appli |
| 39 | 100 | 3.3 | 11283 | 3 | US-09-099-753-3 | Sequence 3, Appli |
| 40 | 100 | 3.3 | 11283 | 3 | US-08-986-106-3 | Sequence 3, Appli |
| 41 | 99.5 | 3.3 | 3159 | 4 | US-09-134-000C-3215 | Sequence 3215, Ap |
| 42 | 99.5 | 3.3 | 4702 | 4 | US-08-956-171E-268 | Sequence 268, App |
| 43 | 99.5 | 3.3 | 5923 | 4 | US-09-976-594-907 | Sequence 907, App |
| 44 | 99.5 | 3.3 | 5952 | 4 | US-09-687-875A-1 | Sequence 1, Appli |
| 45 | 99.5 | 3.3 | 7886 | 2 | US-08-751-189-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1

US-08-811-583-1
; Sequence 1, Application US/0811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 3731 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Tomato
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 194..3535
 US-08-811-583-1

Alignment Scores:

Pred. No.: 2,266-20 Length: 3731
 Score: 262.00 Matches: 127
 Percent Similarity: 37.66% Conservativity: 85
 Best Local Similarity: 22.56% Mismatches: 222
 Query Match: 8.70% Indels: 130
 DB: 3 Gaps: 17

US-09-913-878a-2_COPY_710_1282 (1-573) x US-08-811-583-1 (1-3731)

QY 1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
 Db 1772 GGAATTGGTAAATATCTCGTGACCTTGTCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831
 QY 21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
 Db 1832 TATACCCCATCTGCTTCCAGATTCGTATGTGTGATATAAGGTGTGTGGGTGTGAT 1891
 QY 41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyProSerGlnArgIleAspTrpGlu 60
 Db 1892 CCGGATTCATCAATGAAG-----TTGTCTTTGAGAAGAGCATCTCGAATATGAA 1942
 QY 61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
 Db 1943 TCACACAACATA-----AAGTTAGATGTCCTTGGGA---TGGAGCAAAATAT 1984
 QY 81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaAspLys 100
 Db 1985 CAGCCTTGTTATCTTAATCGTCACTGATGATAGCTCTGTCT----- 2026
 QY 101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
 Db 2027 -----ACACTTGGAGTGAAGATGAATGAACTTCTCGAACAGACAAAAGGAAGCTGTA 2077
 QY 121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyGluSerTy 140
 Db 2078 GATCAGCTTGATGCTATCTTGCATGATCTTTGAAGGCACAGAGCGCTTTGGAATTGATG 2137
 QY 141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
 Db 2138 TCT-----CCTGGA 2146
 QY 161 SerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
 Db 2147 GAGAACACTAAATATCTCAAGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206
 QY 181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
 Db 2207 TTTCTTTCAATGATGTTGGCAACCTTCGCGCATCCAAAGTTCTCGATTTGGGGACTAGA 2266
 QY 201 LeuAsnIleArgValGlyArgSerAlaTyIleTyrMetIleAlaAspPheTrpGlyVal 220
 Db 2267 TCAAGATATTTATTCCAATGGAAGACAATGATGGGATGTTTGGATCAATCCAGAAC 2326
 QY 221 LeuGluGluAsnGluValHisValGlyPheSer-----SerLysPheArgAsp 236
 Db 2327 TTGGAATATGGTCAGGTGTTGTTCAGTTTACTGTGTGTCGACATGGAGAGTTTCTGAC 2386
 QY 237 GluGluGluSerPhe-----ThrLeuLeuSerAspCysAsp 248

Db 2387 GATTTCATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAT 2446
 QY 249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
 Db 2447 GTGGTTGTCGCAAAATCCATGCTTGCATCTCTGGTGATATTCTGTTTAAAGGCTGTA 2506
 QY 269 PheLysProGluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspVal 288
 Db 2507 AATGTTTCGAGCGCTGCACCATAGGTAGATTGTGTGTATTCCTCAGAAAGAAAAAGA 2566
 QY 289 ProLeuAlaLysLysLeuSerGlyGlyAspTyArgAspGlyAspMetAlaTrpValCysTrp 308
 Db 2567 CCTCATCCGAATGAATGTTCTGGAGTGATTGGATGGGATATCTACTTTGTTTCTGCG 2626
 QY 309 AspProGluLeuValAspGlyPheValAsnAlaGluMetProLeuGlnProAspLeuSer 328
 Db 2627 GATCAAGACATGATCCCGCAAGGCAAGTCCAGCCGATGGAATATCTCCAGCACCAGC 2686
 QY 329 ArgTyLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
 Db 2687 ATACAGTTGACCATGAT----- 2704
 QY 349 SerAlaAlaLysGluGlnThrThrTyArgMetIleGlnLysSerPheHisPheAlaLeu 368
 Db 2705 -----GTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATT 2746
 QY 369 GlnProAsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCys 384
 Db 2747 GTGAATGACAGTTTGGGAATCATAGCAATGCCATGCTCGTATTTCACACAGA----- 2800
 QY 385 TyrIleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsn 404
 Db 2801 --GAACCTGATATGTCATGAGTATGATCCATGCAAAAGAACTTCTGAGCTCTTTTCAATT 2857
 QY 405 LeuValAsnGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
 Db 2858 GCAGTGGACTTCCAAAGACTGTGTGTTCGCGTGAATACCATCT-----CAGTTGCGC 2911
 QY 425 ArgGluLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
 Db 2912 -----CCTAAAGATACCCAGACTTCTCATG 2935
 QY 444 PLeuGlyArgGlyGluProThrHisIleIleAspTyLeuLysPheSerIleAlaArgPr 464
 Db 2936 GATAAGCCGG-----ACAAGACC 2953
 QY 464 oAlaIleAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
 Db 2954 AGCTATATCTCAGAAAGAGTTATTGAAAGCTTTTC-AGGAAAGTGAAG----- 3001
 QY 482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyThrPhePheLys 502
 Db 3002 -----GACAAAGCA-----CCTCAGGCTAGTCTCTATCGCAGCTTTCAGAG 3042
 QY 502 sGluIleSerAspLysSerArgSerSerAlaLeu----- 513
 Db 3043 AGATGTTGCAAGGAGATCATATGATCTGATATGGAAGTTGATGGATTGAAGATTACAT 3102
 QY 514 -----LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGln 528
 Db 3103 TGACGAAGCTTTTGACTACAAAACCTGAATATGACAAACAGCTGGGTAAATTAAATGGA 3162
 QY 528 uTyGly 530
 Db 3163 CTATGGC 3169

RESULT 2

US-09-913-294A-5730
 ; Sequence 5730, Application US/09313294A
 ; Patent No. 6476212
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath V.

; APPLICANT: Ito, Laura Y.
 ; APPLICANT: Sherman, Bradley K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
 ; FILE REFERENCE: PL-0017 US
 ; CURRENT APPLICATION NUMBER: US/09/313,294A
 ; CURRENT FILING DATE: 1999-05-14
 ; NUMBER OF SEQ ID NOS: 7600
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 5730
 ; LENGTH: 274
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6476212 700350587H1
 ; NAME/KEY: unsure
 ; LOCATION: 41, 232, 244, 248, 257-258, 268-269
 ; OTHER INFORMATION: a, t, c, g, or other
 ; US-09-313-294A-5730

Alignment Scores:

| Pred. No.: | 8, 6e-06 | Length: | 274 |
|------------------------|----------|---------------|-----|
| Score: | 124.00 | Matches: | 24 |
| Percent Similarity: | 56.16% | Conservative: | 17 |
| Best Local Similarity: | 32.88% | Mismatches: | 30 |
| Query Match: | 4.12% | Indels: | 2 |
| DB: | 4 | Gaps: | 1 |

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-313-294A-5730 (1-274)

RESULT 3

; Sequence 60, Application US/09484970B
 ; Patent No. 6426196
 ; GENERAL INFORMATION:
 ; APPLICANT: Jones, Karen A.
 ; APPLICANT: Volkmut, Wayne
 ; APPLICANT: Walker, Michael G.
 ; TITLE OF INVENTION: BONE REMODELING GENES
 ; FILE REFERENCE: PB-0014 US
 ; CURRENT APPLICATION NUMBER: US/09/484,970B
 ; CURRENT FILING DATE: 2000-01-18
 ; NUMBER OF SEQ ID NOS: 172
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 60
 ; LENGTH: 13977
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
 ; NAME/KEY: unsure
 ; LOCATION: 11721-11761, 12294, 13969
 ; OTHER INFORMATION: a, t, c, g, or other
 ; US-09-484-970B-60

Alignment Scores:

| Pred. No.: | 0.0316 | Length: | 13977 |
|------------------------|--------|---------------|-------|
| Score: | 117.50 | Matches: | 130 |
| Percent Similarity: | 32.36% | Conservative: | 102 |
| Best Local Similarity: | 18.13% | Mismatches: | 254 |
| Query Match: | 3.90% | Indels: | 231 |
| DB: | 4 | Gaps: | 33 |

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-484-970B-60 (1-13977)

QY 24 SerAlaValGlnGlyArgPheGlySer-----AlaLysGlyMetTrpValIleAspVal 41
 DB 1376 ACAGCCCATCAGGCGCGGTGGTAAATATTACAAATGCGAAGTAAGCTGATGGAACA 1435
 QY 42 AspAspThrGlyAspGluAspTrpIleGluThrTrpProSerGlnArg-----57
 DB 1436 GGAAAAATTATCAGAGATCAGAAACTGAAGTACAGACAGACAGATGATCTCCTAAATTC 1495
 QY 58 LysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAla 77
 DB 1496 AGATGGGAATGCCTCAGGCTAGCATGCAAGAAACAAAGCAATTTCATAGAGTTTA 1555
 QY 78 SerGluLeuLysSerAlaGlyLeuAsn-----LeuGlnLeuLeuProValLeuGluAsp 95
 DB 1556 ATGGATCTCAGAACTCAGAACTGAAGAGTTGAATGCTGCTACAAAAACAGAGAA 1615
 QY 96 ArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeu 115
 DB 1616 AGAACCAAGG-----AAATGGAGGAGAGCCTCTTGGACCTGATCTTGAAGACCTA 1666
 QY 116 GlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln--- 134
 DB 1667 AAACGCAAGTACAAACATAG--GTGCTCAAGAGATCTAGAACAGAACAGTC 1723
 QY 135 -----TrpValTyrGluSerTyrSerSerArgAla 144
 DB 1724 AGGTCATTTCTCACTCACTGCGGTGGTAGTTGATGAATCTAGTGAGATCACGCA 1783
 QY 145 ThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGlu 164
 DB 1784 ACTGCT-----GCTTTGGAAGAA 1801
 QY 165 ThrLeuAsnPheLeuMetAsn-----SerGlyPheAspProLysLysGln 179
 DB 1802 CAACTTAAGGTATGGAGATCGATGGGCAACATCTGTAGATGGACAGACCGCTCG 1861
 QY 180 LysTyrLeuGlnAspIle-----AlaTrpAsp---LeuGlnLysArgLysCysAspThr 196
 DB 1862 GTTCTTTTACAAGACATCTCTCAATGGCAACGCTTACTGAAGAACAGTGC----- 1915
 QY 197 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 216
 DB 1916 -----CTTTTACT 1924
 QY 217 PheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAsp 236
 DB 1925 GCATGGCTTTCAGAAAAAAGAGATCAGTGAACAAGATTACACAACTGGCTTTAAAGAT 1984
 QY 237 GluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAla 256
 DB 1985 CAAATGAATGTTTCAAGTCTTCAAAACTGGCGGTTTTAAAGCG-----2032
 QY 257 HisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeu 276
 DB 2033 -----GATCTAGAAAAAAGAAAGCAATCCATG---GGCAAACTGATTTCACCTC 2077
 QY 277 LysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGly 296
 DB 2078 AAACAAGATCTTCTT---TCAACATGAAGAATAAGTCAGTCAGCCAGAGACGGAA--- 2131
 QY 297 GlyAspTyrAspGlyAspMetAlaTrpVal-----CysTrpAsp-----309
 DB 2132 -----GCAATGGCTGGATAACTTTTCCCGGTGTTGGGATAATTTA 2170

QY 310 -----ProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluPro 325
Db 2171 GTCCAAAACTTTGAAAGAGTACACACAGATTTTCCACAGGCTGTCCACACACACACGCA 2230
QY 326 AspLeuSerArgTyrLeuLysAspLys-----ThrThrPheLysGlnLeu 341
Db 2231 TCACATAACACACACACTGTAATGGAACAGTAACAGTACAGTCCACCAAGGGAACAGATC 2290
QY 342 MetAlaSerHisGlyThrGlySer-----AlaAlaLysGluGlnThr 355
Db 2291 CTGGTAAGCATGCTCAAGAGAACTTCCACACACCTCCCAAGAGAGAGGACAGATT 2350
QY 356 ThrTyrAsp-----Met 359
Db 2351 ACTGGGATCTGAATAGGAAGAGTTGGATGTGATATTAACCTGAACCTCAGACGTTG 2410
QY 360 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeu-----Gly 374
Db 2411 ATTACTCGCTCAAGAGCTGTGTTCAGAGTCTCGAATTTGCAATCTTCGGAAGGAAGGC 2470
QY 375 MetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLys--- 393
Db 2471 AACTTCTCAGACTTAAAGAAAAGTCAATGCCATAGAGCGAGAAAAGCTGAGAAGTTC 2530
QY 394 -----ProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGln 408
Db 2531 AGAAAACTGCAAGATGCCAGCAGATCAGCTCAGCGCCTGGTGAACACAGATGTTGAATGAG 2590
QY 409 SerLysGlnGlyLeuValPheAsnGluAlaSer----- 419
Db 2591 GGTGTTAATGCAGATAGCATCAAAACAGCCTCAGAACCACTGAACAGCGGTGATCGAA 2650
QY 420 -----TTPAlaGlnLeuArgArgGluLeuLeu--- 428
Db 2651 TTCTGCCAGTTCGTAAGTCAGAGACTTAACTGGCTGGAGTATCAGAACACATCATCGCT 2710
QY 429 -----GlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTripleu--- 445
Db 2711 TTCTATAATCAGCTCAACAATTTGGAGCAGATGACAACCTACTGCTGGAACCTGGTTGAA 2770
QY 446 -----GlyArgGlyGluProThrHisIleAspTyrLeuLysPheSerIle 461
Db 2771 ATCCAAACCCACCACCCCATCAGAGCCACAGCAATTAAGTCAGTTAAAT---ATT 2824
QY 462 AlaArgProAlaIleAspLys-----GluLeuGluAlaPheHisAsn 475
Db 2825 TGTAGGATGAAGTCAACCGGTATCAGGCTCTCAACCTCAAAATTTGAACGATTAATAAT 2884
QY 476 AlaMetLysAlaAlaLysAspThrGluAspGlyValaHisPheTrpAspProAspLeuAla 495
Db 2885 CAAAGCATAGCCCTGAAAGAGAAAGACAGACAGACCCATGCTTCTGATGCGAGCTTTGTG 2944
QY 496 SerTyrTyrThrPhePheLysGluIle---SerAsp-----LysSerArgSerSer 511
Db 2945 GCCTTTACAATCATTTTAAGCAAGTCTTTTCTGATGTGTCAGGCCAGAGAAAGAGCTA 3004
QY 512 AlaLeuLeuPheThrThrLeu----- 518
Db 3005 CAGACAATTTTGACACTTTGCCACCAATGCGCTATCAGGACACCATGAGTCCCATCAGG 3064
QY 518 ----- 518
Db 3065 ACATGGTCCAGCAGTCAGAAACCAAACTCTCCATACCTCAACTAGTGTCCACCGACTAT 3124
QY 519 -----LysAsnArgIleGlyGluVal-----GluLys 527
Db 3125 GAAATCATGAGCAGAGACTCGGGAAATTGCGGGTTTACAAGTTCTCTCGAAGAGCAA 3184
QY 528 GluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLysAspProTyrProVal 547
Db 3185 CAAAGTGGCCTATATCTCAGCACCATCTGTGAAGAGATGTGGAAGAGAGCGCCCTCT 3244
QY 548 ArgValAsnGlnValTyrGlu-----LysTrp----- 556

Db 3245 GAAATTAGCCGGAATATCATCAAGAAATTGAAGAAATTGAGGAGCGCTGGAGAGCTC 3304
QY 557 -----CysAlaIleThrProGluAlaMetAspLys 566
Db 3305 TCCTCCAGCTGTTGAGCATTTGTCAAAGCTAGAGAGCAAAATGAATAA 3355
RESULT 4
US-09-543-681A-2293
; Sequence 2293, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRAB.
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709, 1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2293
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2293
Alignment Scores:
Pred. No.: 0.00213 Length: 1491
Score: 114.00 Matches: 88
Percent Similarity: 38.18% Conservative: 80
Best Local Similarity: 20.00% Mismatches: 152
Query Match: 3.75% Indels: 120
Gaps: 22
US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-543-681A-2293 (1-1491)
QY 168 PheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrp 187
Db 223 TACATCTTGAAAAAATAAATAAAGATAGCTGCTGAGAAATTAAGATATT----- 276
QY 188 AspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArg 207
Db 277 ---CTTCAAAAAGAAAG-----TTAACTGATAAAATAAAATCTTGATAATAAG 324
QY 208 SerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluLysGlnValHis 227
Db 325 -----GTAATTGTATCA-----GGAATAAAGATGCTAATGATATAAT 363
QY 228 ValGlyPheSerSerLysPheArgAspGluGluSerPheThrLeuLeuSerAsp--- 246
Db 364 GTATTATTCTCTTTGTTAAGATAAAGATGATTCTAGTACACTTTTAAATAAATA 423
QY 247 -----CysAspValLeuValAlaArgSerProAlaHisPhePro 259
Db 424 TTTAGTAATAAAATACTATAATTAATAGTTTCTCTTAAATAAAGAAAGATTCTT 483
QY 260 SerAspIleGlnArgVal-----ArgAlaValPheLysProGlu----- 272
Db 484 GATGACATAGAAATATTGCAATAAATAATTCGTAACCTAAATTTACATTATATAAT 543
QY 273 ---LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 291
Db 544 AATGTAATAAATAATCAATAATAATAATTAATTAATTCGAATAA-----ATTTT 591
QY 292 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 311
Db 592 AAGGAACCAAGAACTAGTATGATATAACAAAGAT-----GATTCTCAT 633
QY 312 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 331
Db 634 AATAAAATAATTTTGAATAAAGAGAA-----AAACAGACCTCTCAAAAAAACC 694

```

QY 332 LysLysAspLysThrThrPhelLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 351
Db 685 AAAAAATCAAAAAA-----CGAGCCCTG 708
QY 352 LysGluGlnThrThrTyAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371
Db 709 ATGGACAACTTGATAAGAT-----AAACAGAAAATGAATGAGCTTTAGCTACTTCT 762
QY 372 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 391
Db 763 ACCCAAGGGTTATTAATAAACTGTAGAG-----TATATAGATAAT---CTATCT 807
QY 392 AsnLysProAlaIleLeuSerSerLeuValGlyAsn-----Leu 405
Db 808 AAT-----TTAATAGTTCTAAACTGATAATGTTACTGTGAATTCGACCTC 855
QY 406 ValAspGlnSerLysGlnGlyIle-----ValPheAsnGlu 417
Db 856 ATTGACCAATTCATAATGCTATTGAACTTTGGTAAAAAATGTAATGCTCTTTAAACAA 915
QY 418 AlaSerTrpAlaGlnLeuArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
Db 916 GATAGTCAGTTTATCTATAAAAAAGATAT-----GAGATAGAGTTAAGTGAC--- 963
QY 438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleAspTyrLeu 457
Db 964 -----ATA 966
QY 458 LysPheSerIleAlaArgProIleAlaLysLysGluLeuGluAlaPheHisAlaMet 477
Db 967 AAAAAACAAT-----GAGAACTTAATGAATAATATATAAAAAAGATG 1008
QY 478 LysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
Db 1009 GATGGTTTAAATCTAATCAGTATATATGTTTATAAGGAAGCAGATAATCTAATTT 1068
QY 498 TyrThr-----PhePheLysGluLysSerAsp 506
Db 1069 ACTCAGTTAACGAGATTAATTTTAAATAAGGTTTTTTTAAATAATATAAAGAT 1128
QY 507 LysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGlu 526
Db 1129 ATGTTTCTCAATATCTCTTATATAATTAATCAATTAAGGAATAATAAGGTTTTTCTCT 1188
QY 527 LysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys---AspProTyr 545
Db 1189 AACTATATAAATTAATTTAAAAACAAATACCTTAAGAAAAAAGAGATATAATGATAAT 1248
QY 546 ProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAsp 565
Db 1249 GCTGTTAAAAAATTAGAAATATTGATAGGAATTTATTCCTTTGCTTGAATCTCTAGAT 1308

```

```

RESULT 5
US-08-961-527-163
Sequence 163, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 163:
SEQUENCE CHARACTERISTICS:
LENGTH: 8494 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-163

Alignment Scores:
Pred. No.: 0.0481
Score: 113.00
Length: 8494
Percent Similarity: 35.13%
Matches: 125
Best Local Similarity: 21.74%
Conservative: 77
Query Match: 3.75%
Mismatches: 216
Indels: 157
Gaps: 28
DB: 4

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-961-527-163 (1-8494)
QY 12 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 31
Db 3555 CGTATCTACGACCTGTTAAATCTGCTGATTAA-----GCTATTCTAT 3596
QY 32 SerAlaLysGlyMetTrpValIleAspValAspThrGlyAspLysLeuLeuGlu 51
Db 3597 AAAGCCAAAGACATTTCTTGTGATGAAGCAGATATGACCTTGATATGGATTCTTGGAA 3656
QY 52 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 71
Db 3657 ACT-----GTTGATAAGATTCTGTCAGCTCT 3683
QY 72 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuPro 91
Db 3684 CCA---AAAGACTTGCAATTCATGCTCTCTCAGCGACTATCCCAAAAACCTGCAACCA 3740
QY 92 ValLeuGlu-----AspArgAlaArgAspLysValLysMetArgGlnAlaIle 107
Db 3741 TTCTTGAAAAAATACTTATCAATCTCTTATGAGAAAATTAAGACCAAAAACGGTTATT 3800
QY 108 GlyAspArgLeuLeuAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAla----- 125
Db 3801 TCTGACACCATTTGATAATGTTGTTGATTTCGACCAAGGACATGATAAGAATGCTCAAAAT 3860
QY 126 -----LeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db 3861 TACCATTGACTCAGTTGATGCGCCG-----TATTTGCAATGATTTTGTAACTAT 3914
QY 141 SerSerArgAlaThrArgValSer----- 148
Db 3915 AAAACGCTGCTGATGATGATTTGATTCATATCTGCTCAAGGCTTGAAGGTTGCAAAA 3974
QY 149 ---HisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGlnThrLeuAsn 167
Db 3975 ATCCATGGCGGATATT-----GCCCTCTGTTGAACGCAAGCAATCATGAAT 4019
QY 168 PheLeuMetAsnSerGlyPheAspProLysLysGlnLysGlnLysGlnAspIleAlaTrp 187
Db 4020 CAGGTGCAAAATCTGGATTTCAG-----TATATT-----GTGCAACA 4058
QY 188 AspLeuGlnLysArgLysCys-----AspThrLeu 197

```


Db 4059 GATTTGGCAGCGGTGGGATTGACATTGAAGGTGTGAGCCATGTCATCAATGATGCCATT 4118
Qy 198 LysSerLysLeuAsnIle-----ArgValGlyArgSerAla----- 209
Db 4119 CCGCAAGACTTATCTTTTGTTCATCGTGTGTGCTGCTACTGGACGAATGGCGCTACCA 4178
Qy 210 -----TyrIleTyrMetIleAlaAspPheTTPGlyValLeuGluAsnGlu 225
Db 4179 GGTACAGCTATTACCTTTATCAGCAAGATGATGACTCGGATCGGTGAGTGGAGAAA 4238
Qy 226 ValHisValGlyPheSerLysPheArgAspGluGluGluSerPheThrLeuLeuSer 245
Db 4239 TTGGGATCAAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 4295
Qy 246 AspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVal 265
Db 4296 GACCGTGATCGTGTGCCCAACCGTGTGAGAAAAAACAAGATAAACTTGCATCGAAATGATT 4355
Qy 266 ArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLys 285
Db 4356 GGTTCGTGTTAAAGAAAAG----- 4376
Qy 286 GlyAspValProLeuAlaLysLysLeuSerGlyAspTyrAspGlyAspMetAlaTTP 305
Db 4377 -----AAAAAGTCAAAACCGGT---TATAAGAGAAAATTCATGG 4415
Qy 306 ValCysTTPAspProGluLeuValAspGlyPheValAsnAlaGluMetProLeuGluPro 325
Db 4416 GCG-----GTTGATGAA----- 4427
Qy 326 AspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHis 345
Db 4428 -----AAGCGCGTAAACCAACCGGTGTGAAATCGCGCTCGCGGT 4469
Qy 346 GlyThrGlySerAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHis 365
Db 4470 CGTGCAGAGCGTAAAGCTAAACGCCAACAA-----TTTTAA 4505
Qy 366 PheAlaLeu-----GlnProAsnPheLeuGlyMetCysThrAsnTyrLysGlu 381
Db 4506 TAGAAATGTTGGAGTATTGAGCTCCCACTTTTATTATGATGAGAACCACTATCTAAC 4565
Qy 382 ArgLeuCysTyrIleAsnAsnSer-ValSerAsnLysProAlaIleLeuSerSerLe 401
Db 4566 CGAAACACTACATTAAAGACTCGMAATTCGATTAAATAATGTTATATGATA----- 4617
Qy 401 uValGlyAsnLeuValAspGlnSerLysGln---GlyIleValPheAsnGluAlaSerTr 420
Db 4618 -----AAGTTATAGTCCCGATAGATGATGATGATGATGATGATGATGATGATGAT 4670
Qy 420 p-----AlaGlnLeuArgGluLeuLeuGlyGlyAlaLeuSerLeuProAsp-- 436
Db 4671 TCAGTACTTTGTAACCTCTATAACAATATTTTAAAGGGGGACATTTTATGTCAGAGCG 4730
Qy 437 -ProMetTyrLysSerAspSerTTPLeuGlyArgGlyGluProThrHisIleIleAspTy 456
Db 4731 TAAATTAATTCAGCTCTGAACT---GTATCTAGGGGCGCATCCGATAGATGATGATGATGAT 4787
Qy 456 rLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAl 476
Db 4788 AATTTCAGATGCGATTGCGATGCTATTTAGCAAGAGATCCAGAGGCG---CAGCTTGC 4844
Qy 476 aMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTTPAspProAspLeuAlaSe 496
Db 4845 TGCTGAAACAGCTGTATATACGTGTTCTGTCACGCTTTTGTGTAATTTCTACAATGC 4904
Qy 496 rTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheTh 516
Db 4905 CTATGTGGATATTAAACCGTGTGTTGCTGAT----- 4935
Qy 516 rThrLeuLysAsnArgIleGlyGluValGluLysGlnTyrGly 530
Db 4936 -ACATTGCA---GAGATGGTATTACCAATACAGAAATATGGA 4974

RESULT 6

US-08-987-439-2
; Sequence 2, Application US/08987439
; Patent No. 6110457
; GENERAL INFORMATION:
; APPLICANT: Beltshe, Robert B.
; APPLICANT: Ray, Ranjit
; TITLE OF INVENTION: LIVE ATTENUATED VACCINES BASED ON CP45 HP1V-3 STRAIN
; TITLE OF INVENTION: AND METHOD TO ENSURE ATTENUATION IN SUCH VACCINES
; FILE REFERENCE: SLU 4481
; CURRENT APPLICATION NUMBER: US/08/987,439
; CURRENT FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6702
; TYPE: DNA
; ORGANISM: Human parainfluenza virus 3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(6702)
US-08-987-439-2
Alignment Scores:
Pred. No.: 0.0375 Length: 6702
Score: 112.50 Matches: 128
Percent Similarity: 35.57% Conservative: 100
Best Local Similarity: 19.97% Mismatches: 214
Query Match: 3.74% Indels: 199
DB: 3 Gaps: 33

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-987-439-2 (1-6702)

Qy 40 AspValAspThrGlyAspGluAspTTPleGluThrTyrProSerGlnArgLysTTP 59
Db 127 GATATGGATGAC-----GACTCAATAGTATTATCACT-----AGACAGAAATA 171
Qy 60 GlucCysAspPheValAspLysHisGlnArgThrLeu----- 71
Db 172 AAATTAATAATTGGATAAAGACACGATCTATTAGAAGATTAAATTAATTAATACT 231
Qy 72 -----GluValArgSerVal 76
Db 232 GAAAAAGTGAATGACTTAGGAAATAACACATTTTATCAGATATCCAGAAATGTCAAAGAA 291
Qy 77 AlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArg 96
Db 292 ATGTTCAATATATATATACCTGGTATTAAACAGTAAAGTGAATTAATTAATAAGCA 351
Qy 97 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsnAspLeuGln 116
Db 352 GATAGAACATATATAGTCAAAATGACTGATGATTAAGAGATCTATGGATTAATGTCTATCA 411
Qy 117 ArgGlnPheSerGlu-----GlnLysHisAlaLeuAsnArgProValGlu----- 131
Db 412 AAATAGCCTCAAAAATGATGGAACCAATTATGATCTTAATGAAGAAATTAATAATA 471
Qy 132 -----PheArgGlnTTPVal 136
Db 472 TCGAAAGTTCACACAACTATAAATCAGATAAATGATATAATCCATCAAAACATGTTT 531
Qy 137 TyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 156
Db 532 ACTATCAAGTATGATGAGAGATGACAAAGCTCGAAATGAGATCACTTTTAATGTT 591
Qy 157 GlyLeuPro---AspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAsp 175
Db 592 GCGAAGGATTATAACTTTGTTAGAAGACCAAGAAATTTCTTATTG-----ATACAT 642
Qy 176 ProLys-----LysGlnLysTyrLeu---GlnAspIleAlaTTPAsp 188
Db 643 CCAGAAATGTTTGTGATATTAGATAAACAACATAATGTTTACTTAATTAATCTCTGAA 702

QY 189 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal----- 205
DB 703 TTATGATTGATGATTGACGATGCGAAGCCGATGGAATATAAGTCGATGCTGAAG 762
QY 206 ----GlyArgSerAlaTyrIleTyrMetIleAla---AspPheTyrGlyValLeuGlu 222
DB 763 TTATGATCCAAATTAACAATCTATGATCAGAAAGGTAATAACCTGTGGGAAGTGATAGAT 822
QY 223 GluAsnGluValHisValGlyPheSerLysPheArgAspGluGluGluSerPheThr 242
DB 823 AAATGTTTCCCAATTATGGGA-----GAAAGACCAATTTGAT 858
QY 243 LeuLeuSerAspCysAspValLeu---ValAlaArgSerProAlaHisPheProSerAsp 261
DB 859 GTGATATGTTATTAGAACCACTTCATTAATTCCTTAATCAACCTCATGATCCT--- 912
QY 262 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal---Ile 280
DB 913 GTTAACAACAATAAGAGGAGCTTTTAAAT-----CATGTGTTATCCGAGATGGAATTA 966
QY 281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
DB 967 ATATTGATCTAGAGATCGATT-----AGGAATTTCTGAGTGTAGATTAC--- 1014
QY 301 GlyAspMetAlaTyrValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
DB 1015 -----ATTGATAAAATTTTAGATATATTTAATAAAGCTACACA 1050
QY 321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrPhe----- 338
DB 1051 ATAGATGAATAGACAGATTTCCTTTTTTTGA-----ACATTTGGGCAT 1098
QY 339 LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys-----Glu 353
DB 1099 CCTCCATTAGAACTAGTATTCAGCAGAAAGGTTAGAAAATATATATATTGGAAA 1158
QY 354 GlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeu 373
DB 1159 CAATTAAATTTGACACTATTAAATGTCATGCTATC----- 1197
QY 374 GlyMetCysThr-----AsnTyrLysGluArgLeuLys----- 384
DB 1198 ---TTCTGTACAAATAAATAAATTAACGATATAGAGAGCGATGCTGGACATGCGCTCT 1254
QY 385 -----TyrIleAsnAsnSerValSerAsnLysProAla 395
DB 1255 GTGACATTACCTGATCATGCACAGAACTCATCAAAATGCTTACGGTCAAACTCTGCG 1314
QY 396 IleIleLeuSerSerLeuValGlyAsnLeuValAsp-----GlnSerLysGlnGlyIle 413
DB 1315 ATATCATATGAA-----AATGCTGTGTATTATTACCAGAGCTTTATAGGAATA 1362
QY 414 ValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeu----- 427
DB 1363 AAATTCATTAATTAATCATAGAGCTCAGTTAGATGAGGATTTCACAAAATTTATATGAAAGAT 1422
QY 428 -----LeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeu 445
DB 1423 AAAGCAATATCTCCAAAAAATCAAAATTTGGACACAGTTTATCTCGCATCTAATTTACTG 1482
QY 446 GlyArgGly----- 448
DB 1483 TACCGTACTAACCGCATCCAAACGAATCAGCAAGATTAGTTGAAGTATTTATAGCAGATAGT 1542
QY 449 -----GluProThrHisIleIleAspTyrLeu----- 457
DB 1543 AAATTGATCTCTCATCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1602
QY 458 LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMet 477
DB 1603 GAATTATATTTCTTATAGTCTTAAAGAAAAGAGATCAAA----- 1644

QY 478 LysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
DB 1645 -----CAGGAAGT----- 1653
QY 498 TyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPhe---Thr 516
DB 1654 ---AGACTCTTTGCAAAATGACATACAAAATGAGAGACTACACAGTTTATCAGAGACA 1710
QY 517 ThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgIleuValLysAsnLys 536
DB 1711 CTACTTGCAAATTAACATAGAAATTTCTTCAAGAAAATGGGATGGTGAAGGAGAGATT 1770
QY 537 GluMetArgAspSerLysAspProTyrProVal-----ArgValAsnGlnVal 552
DB 1771 GAATTACTTAAGAGATAACCAACCATCATCAATACAGGATTCACCGGTATATGAAGTG 1830
QY 553 Tyr 553
DB 1831 TAC 1833
RESULT 7
US-08-750-717-1
; Sequence 1, Application US/08750717
; Patent No. 6180109
; GENERAL INFORMATION:
; APPLICANT: MOORMANN, Robertus J. M.
; APPLICANT: VAN RIJN, Petrus A.
; TITLE OF INVENTION: Nucleotide Sequences of Pestivirus
; TITLE OF INVENTION: Strains, Polypeptides Encoded by These Sequences and Use
; TITLE OF INVENTION: Thereof for Diagnosis and Prevention of Pestivirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,717
; FILING DATE: 24-DEC-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94201743.5
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/NL95/00214
; FILING DATE: 16-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: BO 39123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2237
; TELEFAX: 703-685-0573
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12311 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 374..12067
US-08-750-717-1

| | | |
|--|--------|---|
| Alignment Scores: | | |
| Pred. No.: | 0.153 | 12311 |
| Score: | 111.00 | Matches: 102 |
| Percent Similarity: | 35.48% | Conservative: 80 |
| Best Local Similarity: | 19.88% | Mismatches: 187 |
| Query Match: | 3.69% | Indels: 144 |
| DB: | 3 | Gaps: 23 |
| US-09-913-878a-2_COPY_710_1282 (1-573) x US-08-750-717-1 (1-12311) | | |
| QY | 67 | HisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsn 86 |
| DB | 10301 | CACAGACAAATGTGTATAACAAGACAATAAGCTCAGTAATGACAGCTACTGTTATCAGG 10360 |
| QY | 87 | LeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAla 106 |
| DB | 10361 | TGGGAAGATTGCGCGTGGTTAGGCCCCAGACA---GACCCAAACAACTTCCACCAAGCA 10417 |
| QY | 107 | IleGlyAspArgLeuIleAsnAsp-----LeuGlnArgGlnPhe 119 |
| DB | 10418 | ATAAGGATAAGATACAGCAAGCAAGCAACCTACAAACCCCGGTTTACATAAGAAATTA 10477 |
| QY | 120 | SerGlnLysHisAlaLeuAsnArgPro-----ValGlu 131 |
| DB | 10478 | ATGGAAGTTTTCACGCATTGAAACGCCGAGTTAGAGTCTCTACGACGCGGTGAA 10537 |
| QY | 132 | Phe-----ArgGln 134 |
| DB | 10538 | TGGGAGAACTGGAGAGAGGATAACAGGAAGGCTGCTGCTGTTTTCGAAAGCAAA 10597 |
| QY | 135 | TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHis----- 149 |
| DB | 10598 | AATATAGGGGAATATTGATTTCAGAGAAATAAAGTCGAAGAGATTATTGACAATCTG 10657 |
| QY | 150 | -----GlyArg-----ValProPheLeuAlaGlyLeuProAspSerGlnGluLthr 165 |
| DB | 10658 | AAAAAGGTAGAAACATTAAATATTATGAACCGGATCCCAAGAAATGAGAGAGGGAC 10717 |
| QY | 166 | LeuAsnPheLeuMetAsnSerGly-----PheAspProLysGlnLysTyrLeuGln 183 |
| DB | 10718 | GTCACAGTACTGGACCGCGGTGATTTTCGTGGACGAGAAACCTAGAGTATACAA 10777 |
| QY | 184 | -----AspIleAlaTrpAspLeuLysArgLysCysAspThrLeuLysSer 199 |
| DB | 10778 | TACCTGAAGCAAAACAAGACTGCCATCACCAAGGTGATATATAAGTGGTGAGCAG 10837 |
| QY | 200 | LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219 |
| DB | 10838 | AAGCCAGTAGTTATACCGGTATGAAGGAAAGACACACCTCTATTCCAAATTTTGACAAA 10897 |
| QY | 220 | ValLeuGluGlu-----AsnGluValHisValGlyPheSerSerLysPhe 234 |
| DB | 10898 | GTAAGAAGGATGGGATCAATTTCAAAATCCAGTGGCGAGTTCGACACTAAGCGG 10957 |
| QY | 235 | ArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuAlaArgSer 254 |
| DB | 10958 | TGGGAC-----ACCAGGTAAACCAAAAGATTGGAGCTGATAAGG--- 10999 |
| QY | 255 | ProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHis 274 |
| DB | 11000 | -----GACATACAAAGTAT-----TATTTCAAGAAAGAAATGGCAC 11035 |
| QY | 275 | Ser-----LeuLysAspValIlePheSerThrLysGly 286 |
| DB | 11036 | AAATTTATTGACACCTGACCGATATGTCAGAGTACCCGCTGATCAGTCTGATGGG 11095 |
| QY | 287 | AspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla----- 304 |
| DB | 11096 | GAAGTATACATAAGGAAGGCAAGAGGCGAGTGGACAACTGACACAAGTCCGGCAAC 11155 |
| QY | 304 | ----- 304 |

RESULT 8
US-08-804-196-1
Sequence 1, Application US/08804196
Patent No. 5874256
GENERAL INFORMATION:
APPLICANT: Bertina, Rogier
APPLICANT: Reitsma, Pieter
TITLE OF INVENTION: A method for diagnosing an increased
TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
TITLE OF INVENTION: thrombosis and kit for use with the same.
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5874256el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,196
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/454,353
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.

| | | |
|----|-------|--|
| DB | 11156 | AGCATGCTAAATGCTCTTAACAATGGTTTACGCTTCTGCGAGGCCACAGGAGTACCTAC 11215 |
| QY | 305 | -----TrpValCysTrpAspProGluIleValAspGly 315 |
| DB | 11216 | AAGAGCTTTTACAGGGTGGCAAAATTCATGTGCGGGGAT-----GATGCG 11263 |
| QY | 316 | PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 335 |
| DB | 11264 | TTCTGTATCACAGAAAGAGCTCTC-----GCTGAGAAA 11296 |
| QY | 336 | ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 355 |
| DB | 11297 | TTTGCAGTAAGGAGCTCCAGATCCCTTATGAAGCTGGGAAGCCCGAGAG---ATCACT 11353 |
| QY | 356 | ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375 |
| DB | 11354 | GAAGGGGACAAAATGAAGTGGCTTACCAATTT-----CATGATATTGAGTTT 11401 |
| QY | 376 | CysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysPro--- 394 |
| DB | 11402 | TGCTCCCATACCAATACAAAGTAAGATCGTCAGATACACTTCTAGTTTACATGCGCGGG 11461 |
| QY | 395 | -----AlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 411 |
| DB | 11462 | AGAAATACACCAACATCTCGCAAGAGTGGCCACGAGGTAGATTCCAGCGTGAAAGG 11521 |
| QY | 412 | Gly-----IleValPheAsnGluAla-----SerTrpAla 421 |
| DB | 11522 | GGTACCATAGCATAGAAAGCAGTAGCATTTAGCTTCTGCTGATGATCTCTGGAAC 11581 |
| QY | 422 | GlnLeuArgArgGlu-----LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyr 439 |
| DB | 11582 | CCCTAATTAGAAGGATCTGCTTACTGGTGTCTATCACTGAAGTGAAGTGAAGCCAGGG 11641 |
| QY | 440 | LysSerAspSerTrpLeuGlyArgGlyGluPro-----ThrHisIleIleAsp 455 |
| DB | 11642 | AGTCAACTACTTACTATTATGAAAGAGAGCCGATATCTGCTTACAAAGGAAGTCAATCGG 11701 |
| QY | 456 | TyrLeuLysPheSerIleAlaArgProAlaIleAspLys 468 |
| DB | 11702 | CACACCTTTTGTATCTTAAGAGAACAGCTTTGAGAAG 11740 |

```

; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-804-196-1

Alignment Scores:
Pred. No.: 0.0681 Length: 6909
Score: 110.50 Matches: 109
Percent Similarity: 37.12% Conservative: 74
Best Local Similarity: 22.11% Mismatches: 199
Query Match: 3.67% Indels: 112
DB: 2 Gaps: 28

US-09-913-878a-2_COPY_710_1282 (1-573) x US-08-804-196-1 (1-6909)

QY 28 GlyArgPhe-----GlySerAlaLysGlyMetTrpValIle-----AspValAsp--- 42
Db 2923 GGGAGATGGCAATTGGCTTCTGAGAAAGGTAGCTATGAATATCAAGATACATGATGAA 2982
QY 43 AspThrGlyAspGluAspTrp---IleGluThrTrpProSerGlnArgLysTrp----- 59
Db 2983 GACACAGCTGTTAAACAATTGGCTGATCAGCCCCCAGATGCGCTCGTTGGGGAGAA 3042
QY 60 GluCysAspPheValAspLys-----HisGlnArgThrLeuGluVal 73
Db 3043 AGCACCCCTTTCGCCAACAGCTCGAAGACGAGAGTGGCCACCAAGTTCTTCTAGAGTT 3102
QY 74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
Db 3103 AGACAT-----AAATCTCTACAAGTAAGA 3126
QY 94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
Db 3127 CAGATGGAGAAAGAGTAGACTGAAGAAAGCCAGTTT-----CTCAATTAG 3174
QY 114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu----- 131
Db 3175 -----ACACGAAAAAAGAAAGAGAGACACACACCATGCTCTTATCTCCGAGG 3228
QY 132 -----PheArgGlnTrpValTyrgluSerTrpSerArgAlaThrArgVal 147
Db 3229 ACCTTTCACCCCTTAAGAAAGTGAAGCCCTACAACACATTTTCAGAAAGA-----AGACTT 3282
QY 148 SerHisGlyArgValProPheLeuAla-----GlyLeuProAspSerGlnGlu 164
Db 3283 AAGCATTCGTGGTGGCTTCATAAATCCAATGAACATCTCTTCCACAGACCTCAATCAG 3342
QY 165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLysGlnLysTyrrLeuGlnAsp 184
Db 3343 ACATTCCCC---TCTATGGATTGGCTGGATAGCCTCACTTCTCCGACCATAATCAGAA 3399
QY 185 IleAlaTrpAspLeuGlnLysArgLysCys-----AspThrLeuLysSer 199
Db 3400 TCTCAAAATGACACTGGTCAGGACAGCTGCTCCAGGCTTTATCAGACAGTGGCCCCA 3459
QY 200 LysLeuAsnIleArgValGlyArgSerAlaTyrrIleTyrrMetIleAlaAspPheTrpGly 219
Db 3460 GAGGAACACTATCAACA-----TTCCCC 3483
QY 220 ValLeuGluGlnAsnGluValHisValGlyPheSerSerLysPheArgAspGluGlu 239
Db 3484 ATTCAAGACCTGATCAATGACACTCTACTTACAGCCCCAGTCACAGATCTCTTCTCCA 3543
QY 240 SerPheThrLeuLeuSerAspCysAspValLeuAlaArgSerProAlaHisPhePro 259

```

```

Db 3544 GAGCTCAGTGAATGCTTGAGTAGAC-----CGAAGTCACAAGTCTCTCCCC 3591
QY 260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
Db 3592 ACAGATATAGTCAATGATCCCTTCTCTCAGAA-----CATGAAGTCTGCAGACA 3642
QY 280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
Db 3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCTCTCTCCAGAACTCAGCCAGACAACCTC 3702
QY 300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
Db 3703 TCTCCAGACCTCAGCCACAGACTCTCTCTCCAGAACTCATTCCAGAGAACTTCCCCA 3762
QY 319 -----AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
Db 3763 GCCCTCGTCAGATGCCCATTTCTCCAGACCTCAGCCAT-----ACA 3804
QY 337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 356
Db 3805 ACCCTTCTCCAGACCTC-----AGCCATACAACTCTTCTTTAGACCTCAGCCAGACAAC 3861
QY 357 Tyr-----AspMetIleGlnLysSerPhe----- 364
Db 3862 CTCTCTCCAGAACTCAGTCAGACAACCTTTCCCCAGCCCTCGGTCCAGATGCCCTTCT 3921
QY 365 -----HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
Db 3922 CCAGACCTCAGCCATACACCTTTCTCTAGACTTC-----AGCCAGACAACCTCTCT 3975
QY 381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle----- 397
Db 3976 CCAGAACTCAGCCATAGCTCTCT-----CCAGAACTCAGTCAGACAAC 4023
QY 398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
Db 4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CATTCTCCAGACCTCAG---CCATACACCT 4079
QY 418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyAlaLeuSerLeuLeuAspPro 437
Db 4080 TTCTTAGACTTCCAGCCAGACAACCTCTCTCCAGAACTCAGTCAACAACAACTTCCCC 4139
QY 438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleLeuAspTyrLeu 457
Db 4140 AGCCCTCGTCAGATGCCCTTCTCCAGACCCCGCCA-----TACAAC 4184
QY 458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
Db 4185 CTTTCTCT-----AGACCTCAGCCAGACAACCTCTC 4217

RESULT 9
US-08-658-340-1
; Sequence 1, Application US/08658340
; Patent No. 5910576
; Patent No. 5910576 5861489
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; APPLICANT: Reitsma, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased
; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo NO. 5910576 5861489el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,340
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/454,353
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELEPHONE: (301) 948-7400
TELEFAX: (301) 948-9751
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-658-340-1

Alignment Scores:
Pred. No.: 0.0681 Length: 6909
Score: 110.50 Matches: 109
Percent Similarity: 37.12% Conservative: 74
Best Local Similarity: 22.11% Mismatches: 199
Query Match: 3.67% Indels: 112
DB: 28 Gaps: 2

US-09-913-878a-2_COPY_710_1282 (1-573) x US-08-658-340-1 (1-6909)

QY 28 GlyArgPhe-----GlySerAlaLysGlyMetTrpValIle-----AspValAsp--- 42
Db 2923 GGGAGATGGCAATTTGGCTTCTCAGAAAGGTAGTATGAATAATCAAGATGACTGATGAA 2982
QY 43 AspThrGlyAspGluAspTrp---IleGluThrTyProSerGlnArgLysTrp----- 59
Db 2983 GACACAGCTGTAAACATTTGGCTGATCAGCCCGCAGATGCTCAGCTGCTGGGAGAA 3042
QY 60 GluCyAspPheValAspLys-----HisGlnArgThrLeuGluVal 73
Db 3043 AGCACCCCTCTTGGCAACAAAGCTGGAAAGACAGAGTGGCCACCCAAAGTTCTCAGATT 3102
QY 74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
Db 3103 AGACAT-----AAATCTCTCAAGTAAGA 3126
QY 94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
Db 3127 CAGGATGGAGGAAGAGTAGACTGAAGAAAGCCAGTTT-----CTCATTAAAG 3174
QY 114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu----- 131
Db 3175 -----ACAGCAAAAAAGAAAGAGAGACACACACCATGCTCTTTATCTCCGAGG 3228
QY 132 -----PheArgGlnTrpValTyArgLysTyProSerArgAlaThrArgVal 147
Db 3229 ACCTTTACCTCTAAGAGTGAAGCTACACACATTTTCAGAAAGA-----AGACTT 3282
QY 148 SerHisGlyArgValProPheLeuAla-----GlyLeuProAspSerGlnGluGlu 164
Db 3283 AAGCANTCGTGTGGTGTCTATAAATCCAAATGAACATCTCTCCACAGACCTCAATCAG 3342
QY 165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLysTyProLeuGlnAsp 184
Db 3343 ACATTGGCC---TCTATGGATTTGGCTGGATAGCTCCTCTCGACCATTAATCAGAT 3399
QY 185 IleAlaTrpAspLeuGlnLysArgLysCys-----AspThrLeuLysSer 199
Db 3400 TCCTCAATGACATGCTCAGGCAAGCTGTCTCCAGGTCTTTATCAGACAGTGCCTCCCA 3459
QY 200 LysLeuAsnIleArgValGlyArgSerAlaTyIleTyMetIleAlaAspPheTrpGly 219

Db 3460 GAGGAACACTATCAAAACA-----TTCCCC 3483
QY 220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
Db 3484 ATTCAAGACCCCTGATCAATGACATCTACTTCCAGACCCAGTCACAGATCTCTCTTCCA 3543
QY 240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
Db 3544 GAGCTCAGTGAATGCTTGAATGAC-----CGAAGTCACAAAGTCTCTCCCC 3591
QY 260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
Db 3592 ACAGATATAAGTCAATGTCCTCTCTCCAGAACTCAGCAGAACTCAGCAGAACTCTC 3642
QY 280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysSerGlyGlyAspTyr 299
Db 3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCTCTCTCCAGAACTCAGCAGAACTCTC 3702
QY 300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
Db 3703 TCTCCAGACCTCAGCCACACAGACTCTCTCTCCAGAACTCTCAGAGAACTTTCCCCA 3762
QY 319 -----AlaGluMetProLeuGluProAspLeuSerArgTyLeuLysLysAspLysThr 336
Db 3763 GCCCTCGGTGACATGCCATTTCTCCAGACCTCAGCCAT-----ACA 3804
QY 337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThr 356
Db 3805 ACCCTTTCTCCAGACCTC---AGCCATACAACCTTTCTTACCTCAGCCTCAGCAGCAAC 3861
QY 357 Tyr-----AspMetIleGlnLysSerPhe----- 364
Db 3862 CTCTCTCCAGAACTCAGTCAGCAAACTTTCCCGACGCTCGGTGATGATGCCCTTTCT 3921
QY 365 -----HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyLys 380
Db 3922 CCAGACCTCAGCCATACAACCTTTCTCTAGACTTC-----AGCCAGCAAACTCTCT 3975
QY 381 GluArgLeuCysTyIleAsnAsnSerValSerAsnLysProAlaIleIle----- 397
Db 3976 CCAGAACTCAGCCTAAGTACTCTCT-----CCAGAACTCAGTCAGACAAAC 4023
QY 398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
Db 4024 CTTTCCCGACCTTGTGTAGATGCC-CATTCTCCAGACCTCAG---CCATACAACTCT 4079
QY 418 AlaSerTrpAlaGlnLeuArgGluLeuGlyGlyAlaLeuSerLeuProAspPro 437
Db 4080 TTCTCTAGACTTCCAGCAAACTCTCTCCAGAACTCAGTCAGTCAAACTTTCCCC 4139
QY 438 MetTyLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
Db 4140 AGCCCTCGGTGATGCCCTTTCTCCAGACCCAGCCA-----TACAAC 4184
QY 458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
Db 4185 CTTTCTCT-----AGACCTCAGCCAGCAAACTCTCTC 4217

RESULT 10
US-08-746-111-26
Sequence 26, Application US/08746111
Patent No. 6068778
GENERAL INFORMATION:
APPLICANT: Ginsburg, David
APPLICANT: Cui, Jisong
TITLE OF INVENTION: Compositions And Methods For Screening
TITLE OF INVENTION: Compounds For Anticoagulant Activity
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco

```

; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-746-111-26

Alignment Scores:
Pred. No.: 0.0681 Length: 6909
Score: 110.50 Matches: 109
Percent Similarity: 27.12% Conservative: 74
Best Local Similarity: 22.11% Mismatches: 199
Query Match: 3.67% Indels: 112
Gaps: 28
DB: 3

US-09-913-878a-2_COPY_710_1282 (1-573) x US-08-746-111-26 (1-6909)

QY 28 GlyArgPhe-----GlySerAlaLysGlyMetTrpValIle-----AspValAsp--- 42
DB 2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAATAATCCAGATACGTATGAA 2982
QY 43 AspThrGlyAspGluAspTrp---IleGluThrTrpSerGlnArgLysTrp----- 59
DB 2983 GACACAGCTGTTAACAAATTGGCTGATCAGCCCCCAGAAATGCTCAGCTGCTGGGGAGAA 3042
QY 60 GluCysAspPheValAspLys-----HisGlnArgThrLeuGluVal 73
DB 3043 AGCACCCCTCTTGCCACACAGCCCTGGAAAGCAGAGTGGCCCAAAAGTTTCTTAGAGTT 3102
QY 74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
DB 3103 AGACAT-----NAATCTCTACAGTAAGA 3126
QY 94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
DB 3127 CAGGATGGAGGAAAGAGTAGACTGAGAAAGAGCCAGTTT-----CTCATTAAG 3174
QY 114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu----- 131
DB 3175 -----ACACGAAAAAAGAAAGAGAGACACACACCATGCTCTTTATCTCCGAGG 3228
QY 132 -----PheArgGlnTrpValTyrGluSerTyrSerArgAlaThrArgVal 147
DB 3229 ACCTTTACCCCTCTAAGAGTGAAGCCTACACACATTTTCAGAAAGA-----AGACTT 3282
QY 148 SerHisGlyArgValProPheLeuAla-----GlyLeuProAspSerGlnGluGlu 164
DB 3283 AAGCATTTGGTGGCTCTCATAAATGAATGAACATCTCTCCACAGACCTCAATCAG 3342
QY 165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
DB 3343 ACATTGCCC---TCATGGATTTGGCTGGATAGCTTCACCTCCTCGACCAATAATCAGAA 3399

```

```

QY 185 IleAlaTrpAspLeuGlnLysArgLysCys-----AspThrLeuLysSer 199
DB 3400 TCCTCAATGACACTGGTCAGCAAGCTGTCTCCAGGTCTTTATCAGACAGTCCCCCA 3459
QY 200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
DB 3460 GAGGAACACTATCAACA-----TTCCCC 3483
QY 220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGlu 239
DB 3484 ATTCAGACCTGATCAATGCACTCTACTCAGACCCCGAGTCAGACTCTCTCTCTCCA 3543
QY 240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
DB 3544 GAGCTCAGTGAATGCTTGAGTATGAC-----CGAAGTCACAAAGTCTTCCCC 3591
QY 260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
DB 3592 ACAGATATAAGTCAATGTCCCTCTCCAGAA-----CATGAGTCTGGCAGACA 3642
QY 280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTrp 299
DB 3643 GTCATCTCTCCAGACTCAGCCAGGTGACCTCTCTCCAGAACTCATTCAGAGAAACCTTCCCA 3702
QY 300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
DB 3703 TCTCCAGACTCAGCCACAGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTCCCA 3762
QY 319 -----AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysAspLysThr 336
DB 3763 GCCTCGGTGATGATGCTTCTCCAGACTCAGCCAT-----ACA 3804
QY 337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGlnThrThr 356
DB 3805 ACCCTTTCTCCAGACTC-----AGCCATACACCTTCTTTAGACCTCAGCCAGACAAAC 3861
QY 357 Tyr-----AspMetIleGlnLysSerPhe----- 364
DB 3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCCGCCCTCGTCAGATGCCCTTTCT 3921
QY 365 -----HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
DB 3922 CCAGACCTCAGCCATACACCTTCTCTAGACTTC-----AGCCAGACAAACCTTCT 3975
QY 381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle----- 397
DB 3976 CCAGAACTCAGCATATGACTCTCT-----CCAGAACTCAGTCAGACAAAC 4023
QY 398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
DB 4024 CTTTCCCGCCCTTGGTCAGATGCC-CATTTTCTCCAGACCTCAG-----CCATACAACT 4079
QY 418 AlaSerTrpAlaGlnLeuArgGluLeuLeuGlyAlaLeuSerLeuProAspPro 437
DB 4080 TTCTCTAGACTTCAGCCAGACAAACCTCTCTCCAGAACTCAGTCAGACAAACCTTCCCC 4139
QY 438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
DB 4140 AGCCCTCGGTGATGCCCTTCTCTCCAGACCCCGACA-----TACAAAC 4184
QY 458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
DB 4185 CTTTCTCT-----AGACTCAGCCAGACAAACCTTCTC 4217

RESULT 11
US-08-454-353A-1
; Sequence 1, Application US/08454353A
; Patent No. 6518016
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; APPLICANT: Reitsma, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased

```

; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 651801661 Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,353A
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-454-353A-1
;
; Alignment Scores:
; Pred. No.: 0.0681 Length: 6909
; Score: 110.50 Matches: 109
; Percent Similarity: 37.12% Conservative: 74
; Best Local Similarity: 22.11% Mismatches: 199
; Query Match: 3.67% Indels: 112
; DB: 4 Gaps: 28
;
; US-09-913-878a-2_COPY_710_1282 (1-573) x US-08-454-353A-1 (1-6909)
;
; QY 28 GlyArgPhe-----GlySerAlaLysGlyMetTrpValLe-----AspValAsp--- 42
; Db 2923 GGAGATGGCATTTGGCTCTTCTGAGAAAGTACCTATGAAATATCCAGACTACTGATGAA 2982
;
; QY 43 AspThrGlyAspGluAspTrp---IleGluThrTyrProSerGlnArgLysTrp----- 59
; Db 2983 GACACAGCTTTTAAACAATTGGCTGATCAGCCCCCAGATGCTCAGTCTGGGGAGAA 3042
;
; QY 60 GluCysAspPheValAspLys-----HisGlnArgThrLeuGluVal 73
; Db 3043 AGCACCCCTCTTCCCAACAGCTGGAAGAGAGAGTGGCCACCCCAAGTTTCTTAGAGTT 3102
;
; QY 74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeu 93
; Db 3103 AGACAT-----AAATCTCTACAGTAAGA 3126
;
; QY 94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsn 113
; Db 3127 CAGGATGGAGAAAGAGTACACTGAAGAAAGAACCCAGTTT-----CTCATTAAAG 3174
;
; QY 114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu----- 131
; Db 3175 -----ACAGAAAAAAGAAAGAGAGCAGACACACACCATGCTCTTTATCTCCGAGG 3228
;
; QY 132 -----PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 147
; Db 3229 ACCTTTACCCCTCTAAGAGTGAAGCCTTCAACACATTTTTCAGAAAGA-----AGACTT 3282
;
; QY 148 SerHisGlyArgValProPheLeuAla-----GlyLeuProAspSerGlnGluGlu 164
; Db 148 SerHisGlyArgValProPheLeuAla-----GlyLeuProAspSerGlnGluGlu 164
;

3283 AAGCATTCGTTGGTGTCTTCATAAATCCAAATGAACATCTCTTCCACAGACCTCAATCAG 3342
165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
3343 ACATTCGCC---TCTATGATTTTGGTGATGGATGAGCTCCTTCTCAGACCAATATCAGAA 3399
185 IleAlaTrpAspLeuGlnLysArgLysCys-----AspThrLeuLysSer 199
3400 TCTCAATGACACTGGTCCAGCAAGCTGTCTCCAGGCTTTTATCAGACAGTGCACCCCA 3459
200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
3460 GAGGAACACTATCAACA-----TTCCCC 3483
220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGlu 239
3484 ATTCAAGACCTTGATCAATGACTCTACTTCCAGACCCAGTCACAGATCTCTTCTTCCA 3543
240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
3544 GAGCTCAGTGAATGCTTGATGATGAC-----CGAAGTCACAAAGTCTTCTTCCCC 3591
260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
3592 ACAGATTAAGTCAATGTCCT 3642
280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyAspTyr 299
3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCTCTCTCCAGAACTCATTCCAGAGAACCTT 3702
300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
3703 TCTCCAGACCTCAGCCAGCAGCTCTCTCTCCAGAACTCATTCCAGAGAACCTTCTCCCA 3762
319 -----AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
3763 GCCTTCGGTCAGATGCCATTTCTCCAGACCTCAGCCAT-----ACA 3804
337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 356
3805 ACCCTTTCTCCAGACCTC---AGCCATACACCTTTCTTTAGACCTCAGCCAGACAAAC 3861
357 Tyr-----AspMetIleGlnLysSerPhe----- 364
3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTCTCCAGCCCTCGGTCCAGATGCCCTTTCT 3921
365 -----HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
3922 CCAGACCTCAGCCATACACCTTTCTCTAGACTTC-----AGCCAGACAAACCTCTCT 3975
381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle----- 397
3976 CCAGAACTCAGCCATATGACTCTCTCT-----CCAGAACTCAGTCAGACAAAC 4023
398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CAATTTCTCCAGACCTCAG---CCATACAAACCT 4079
418 AlaserTrpAlaGlnLeuArgGluLeuLeuGlyIleAlaLeuSerLeuProAspPro 437
4080 TTCTTAGACTTCCAGCCAGACAAACCTCTCTCCAGAACTCAGTCACAAACCTTTCCCC 4139
438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleLeuAspTyrLeu 457
4140 AGCCCTCGGTGATGCTCTCTCTCCAGACCCCGCCA-----TACAAAC 4184
458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
4185 CCCTTCTCT-----AGACCTCAGCCAGACAAACCTCTC 4217

RESULT 12
US-09-165-019-1
; Sequence 1, Application US/09165019

```

; Patent No. 6558913
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; TITLE OF INVENTION: A method for diagnosing an increased
; risk for thrombosis or a genetic defect causing
; thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/165,019
; FILING DATE:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/454,353
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; US-09-165-019-1

Alignment Scores:
Pred. No.: 0.0681 Length: 6909
Score: 110.50 Matches: 109
Percent Similarity: 37.12% Conservative: 74
Best Local Similarity: 22.11% Mismatches: 199
Query Match: 3.67% Indels: 112
DB: 4 Gaps: 28

US-09-913-878a-2_copy_710_1282 (1-573) x US-09-165-019-1 (1-6909)

QY 28 GlyArgPhe-----GlySerAlaLysGlyMetTrpValIle-----AspValAsp--- 42
DB 2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAATATCCAGATGACTGATGAA 2982
QY 43 AspThrGlyAspGluAspTrp---IleGluThrTrpProSerGlnArgLysTrp----- 59
DB 2983 GACACAGCTGTTAACAATTTGGCTGATCAGCCCCCAGGAATGCTCAGCTGGCTTGGGGAGAA 3042
QY 60 GluCysAspPheValAspLys-----HisGlnArgThrLeuGluVal 73
DB 3043 AGCACCCCTCTGCCACACAGCCTGGAAAGCAGAGTGGCCACCCCAAGTTTCTAGAGTT 3102
QY 74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuLeuProValLeu 93
DB 3103 AGACAT-----AAATCTCTACAAAGTAAGA 3126
QY 94 GluAspArgAlaArgAspLysValIysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
DB 3127 CAGGATGAGGAAGAGTAGACTGAAGAAAGCCAGTTT-----CTCATTAAG 3174
QY 114 AspLeuGlnArgGlnPheSerGluGlnIysHisAlaLeuAsnArgProValGlu----- 131
DB 3175 -----ACACGAAAAAAGAAAAAGAGAGACACACACCATGCTCTTATCTCCGAGG 3228
QY 132 -----PheArgGlnTrpValTyGluSerTySerArgAlaThrArgVal 147
DB 3229 ACCTTTCACCTCTTAAGAAGTGAAGCCTTACAACACATTTTCAGAAAGA-----AGACTT 3282
QY 148 SerHisGlyArgValProPheLeuAla-----GlyLeuProAspSerGlnGluGlu 164
DB 3283 AAGCATTCGTTGGTCTTCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 3342
QY 165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLysTyLeuGlnAsp 184
DB 3343 ACATTGCC-----TCTATGGATTTGGCTGGATGCTCCTCCTGACCATCAATCAATCAAT 3399

```

RESULT 13

```

US-09-620-312D-985
; Sequence 985, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod

```

```

QY 185 IleAlaTrpAspLeuGlnLysArgLysCys-----AspThrLeuLysSer 199
DB 3400 TCCTCAATGACACTCGTTCAGGCAAGCTGTCTCCAGGTCTTTATCAGACAGTGCCTCCCA 3459
QY 200 LysLeuAsnIleArgValGlyArgSerAlaTyIleTyMetIleAlaAspPheTrpGly 219
DB 3460 GAGGAACACTATCAACA-----TTCCCC 3483
QY 220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
DB 3484 ATTCAGACCTGATCAATGCACTCTACTTCAGACCCAGTCACAGATCTCTTCTTCTCCA 3543
QY 240 SerPheThrLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
DB 3544 GAGCTCAGTCAAAATGCTTGATGATGAC-----CGAAGTCACCAAGTCTTCCCC 3591
QY 260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
DB 3592 ACAGATATAAGTCAAAATGCTTCTCTCCAGAA-----CATGAAGTCTGGCAGACA 3642
QY 280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
DB 3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCTCTCTCCAGAACTCAGCCAGACAAACCTC 3702
QY 300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
DB 3703 TCTCCAGACCTCAGCCACACGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTCCCCA 3762
QY 319 -----AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
DB 3763 GCCTCGGTGAGTGCCTATTTCTCCAGACCTCAGCCAT-----ACA 3804
QY 337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGlnGlnThr 356
DB 3805 ACCCTTTCTCCAGACCTC-----AGCCATACAACTTCTTTAGACCTCAGCCAGACAAAC 3861
QY 357 Tyr-----AspMetIleGlnLysSerPhe----- 364
DB 3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCTCCAGCCCTCGGTTCAGATGCCCTTTCT 3921
QY 365 -----HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
DB 3922 CCAGACCTCAGCATACAACTTCTCTAGAGTTT-----AGCCAGACAAACCTCTCT 3975
QY 381 GluArgLeuCysTyIleAsnAsnSerValSerAsnLysProAlaIleIle----- 397
DB 3976 CCAGAACTCAGCCATGACTCTCTCT-----CCAGAACTCAGTCAGACAAAC 4023
QY 398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
DB 4024 CTTTCCCAAGCCCTGGTTCAGATGCC-CATTTTCTCCAGACCTCAG-----CCATACAACT 4079
QY 418 AlaSerTrpAlaGlnLeuArgGluLeuLeuGlyAlaLeuSerLeuProAspPro 437
DB 4080 TTCTCTAGACTTCAGCCAGACAACTCTCTCCAGAACTCAGTCACAAACCTTTCCTCC 4139
QY 438 MetTyLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleAspTyrLeu 457
DB 4140 AGCCCTCGGTTCAGATGCCCTTCTCCAGACCCAGCCA-----TACAAC 4184
QY 458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
DB 4185 CTTTCTCT-----AGACCTCAGCCAGACAAACCTCTC 4217

```


; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yunging
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John, Tillinghast
 ; APPLICANT: Dmanac, Radje T.
 ; TITLE OF INVENTION: No. 658662el Nucleic Acids and
 ; FILE REFERENCE: 784CIP2B
 ; CURRENT APPLICATION NUMBER: US/09/620,312D
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pt_FL_genes Version 1.0
 ; SEQ ID NO 985
 ; LENGTH: 2077
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (287)..(1912)
 ; US-09-620-312D-985

Alignment Scores:
 Pred. No.: 0.0218 Length: 2077
 Score: 107.50 Matches: 107
 Percent Similarity: 31.75% Conservative: 67
 Best Local Similarity: 19.53% Mismatches: 177
 Query Match: 3.57% Indels: 197
 DB: 4 Gaps: 26

US-09-913-878a-2_COPY_710_1282 (1-573) x US-09-620-312D-985 (1-2077)

Qy 19 LeuGlyAspValProSerAlaValGlnGly-----ArgPheGlySerAlaLysGly 35
 Db 572 TTGGGTGACCCAGAGAACCGTTTAAACAGGATTTCTCGAGAGGGGATCT-----622
 Qy 36 MetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyr-----53
 Db 623 -----GATCCAGAACCACTGGGATTCAAATCTGGAGTGAAGTTTTCAGTGG 670
 Qy 54 -----ProSerGlnArgLysTrpGluCysAspPheValAspLysHis-----67
 Db 671 GAGAGCCAGGTGGGAAGAGTTGTCAGTTGTTCTGATGATGATACCCAGGGGGCATTTGAC 730
 Qy 68 ---GlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsn 86
 Db 731 AGCCAGTCAACTGTGAAGACTGTGTACCATCTTTGCTCTAAGCACTATGACTAGTTCT 790
 Qy 87 LeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAla 106
 Db 791 GTTCAGATTATATTTATCTCAGAACATTCAGAGAGATGATCTTCAACAGCTGCAGCTC 850
 Qy 107 IleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeu 126
 Db 851 -----TTCACAGAA-----859
 Qy 127 AsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArg 146
 Db 859 -----859
 Qy 147 ValSerHisGlyArgVal-----ProPheLeuAlaGlyLeu 158

Db 860 -----TACGGTCGCTCGCAATGCGATGAAATTTTCCAAAAGCCTTTC-----901
 Qy 159 ProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLys 178
 Db 902 -----CAGACACTGATGCTTTTGGTTAGACATTGGAGATTTCCCTTATGAA 946
 Qy 179 GlnLysTyr---LeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeu 197
 Db 947 TATAGCTATGGAATCCCAAGGAGGAATGGCATTTTGGATAAGCGT-----991
 Qy 198 LysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPhe 217
 Db 992 -----TTACAGGTGAAGAACATCAACATGAGAAATTCAGATGTT-----1033
 Qy 218 TrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGlu 237
 Db 1034 -----CGAATCACTCATCTGTTCTCTCC-----1060
 Qy 238 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
 Db 1061 GATGTCACCTGCTTCTTACCATCCAGGACTCCAGGTGGCCACACCCCT---GAC 1117
 Qy 258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
 Db 1118 TTTCATGGAAATTTAAAGATATTTGCTGTCATTTCAAGAGCAGATTACAGGCACATGATA 1177
 Qy 278 AspValIleIlePheSerThrLys-----285
 Db 1178 CCGTATGTTAAACCCATCTAAGTTAATGGAAGAGAGATCAATGGCTCAAGGTCACC 1237
 Qy 286 -----GlyAspVal 288
 Db 1238 TGTCGGGACTACTGGAGTATTTTAAAGCATATATTTAAATTTTCAAGAGAGATCTG 1297
 Qy 289 Pro-----LeuAlaLysLysLeu 294
 Db 1298 CCTCACCCCAAGTCCATGCTTACGCCCACTGCTGAAGCCCTACAACTAGCAGCTGCAGCC 1357
 Qy 295 SerGlyGlyAsp---TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
 Db 1358 TCTCCCAAGGACATTTATTTATACACATGGAAGAGTTTGT-----1399
 Qy 314 AspGlyPheValAsnAlaGluMetPro---LeuGluProAspLeuSerArgTyrLeuLys 332
 Db 1400 -----GGGGAGAGAGAACCTTATTGTTCTCCAGACATT-----CTAGAG 1438
 Qy 333 LysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys 352
 Db 1439 GAGAGACCTGTGNAATTCAAACAACTGCTCTGGACCAT-----1477
 Qy 353 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 372
 Db 1478 TTTAAGAGACCAAGAGATGGGTGGGAAGGATTTTCAGCTTTCTACCAAGAGAGCTG 1537
 Qy 373 LeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer---391
 Db 1538 GAGAGAGAAATCAAGGATATATATGAGAACTTCTGCAAGCACATGGTAGCAAGAGCTC 1597
 Qy 392 -----AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAsp 407
 Db 1598 TTCAGACCTTCGAAACCCCTGCAGTGTCTTCACGGGCACTGTAGCTTTGTACATAGCC 1657
 Qy 408 GlnSerLysGlnGlyIleValPheAsnGluAla-----418
 Db 1658 TCAGGCCTCACTGCTTCATAGGTCCTGAGTTGTAGCCCGCAGTTGTTCAACTGATGGTT 1717
 Qy 419 -----SerTrpAlaGlnLeuArg-----Arg 425
 Db 1718 GGACTACTGTTAATAGCACCTCCCTCACCTGGGGGTACATCAGGTATTCTGCTCAATACGT 1777
 Qy 426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspPro-----MetTyrLysSerAspSer 443

```
Db 1778 GAG---CTGGGCGAGCTATTGATTTTGGTGGCGCATATGTGTGGAGCAGGCTTCTTCT 1834
QY 444 TrpLeuGlyArgGlyGluProThrHisIleAlaAspTyrLeuLysPheSerIleAlaArg 463
Db 1835 CATATCGGTAATCCACTCAGCCACTGTGAGGATGCGATT-----GTTGGAAGA 1885
QY 464 ProAlaIleAspLysGluLeuGlu 471
Db 1886 CCATCCATGGATAAAAAAGCTCAA 1909

RESULT 14
US-09-489-039A-4533
; Sequence 4533 Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27/17,747
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4533
; LENGTH: 5307
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4533

Alignment Scores:
Pred. No.: 0.101 Length: 5307
Score: 107.50 Matches: 131
Percent Similarity: 33.48% Conservative: 98
Best Local Similarity: 19.15% Mismatches: 236
Query Match: 3.57% Indels: 219
DB: 4 Gaps: 33

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-489-039A-4533 (1-5307)
QY 8 SerValAlaIleArgIleArgAspValLeuGlyLeuGlyAspVal----- 22
Db 3043 ACGTAGCTAACAAGTCGATATACCCCTGAATGATGGCAGTAGTAAACAGATGCA 3102
QY 23 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 41
Db 3103 TCTCCCACTCTGGAAGGATATAGAGCAGACTCTACGTTGCCCTACTATTACAGAGT 3162
QY 42 AspAspThrGlyAspGluAspTrpIleGluThrTyr-----ProSer 55
Db 3163 GCTCTTATGCGATGAGAACCTGGTTAGTTGAGTACGTTGAAAATTTGGTAAGAATAAT 3222
QY 56 GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSer 75
Db 3223 GAAATAGATTGGATTTTGTATTATGTTCTTCGCTCTAGTAATTCAGTAATGCTACATCT 3282
QY 76 ValAlaSerGluLeuLysSerAlaGlyLeuAsn-----LeuGlnLeu 89
Db 3283 GTTTTATCTCTGTGGCCACAGGCTTCCCTAATAAAGTAGTAAAGCAGCCCTTCTCTTTG 3342
QY 90 LeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAsp 109
Db 3343 CTTAAACCTGCTGATTATTATACCATTTAGATTGATTCGATGACACAGAATGGGTGGA 3402
QY 110 ArgLeuIle-----AsnAspLeuGlnArgGlnPheSerGluGln 122
Db 3403 AATGAGATCCATTTCTTGGTCTTAAATAGGATGTCATTGTCGAAAATATATACCTAGAAG 3462
QY 123 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 142
Db 3463 AGA-----AGAGAAGCAGCACTTCGCCATCGAGAAAGAGTCAATTAGAAACC 3510
QY 143 ArgAlaThrArgValSerHisGlyArgValPropLeuAlaGlyLeuProAsp----- 160
```

```
Db 3511 TTATTAACAAGGCTTCAG-----TTCGTAAATGAGTTAAGGATGATATT 3555
QY 161 -----SerGlnGluGluThrLeuAsn 167
Db 3556 TTAATAATTTGATGAACTGAATAATGAAGCTACTCCAAAGTAAATGAATAAGCCITCGG 3615
QY 168 PheLeuMetAsnSerGlyPheAspProLysGlyGlnLysTyrLeuGlnAsp----- 184
Db 3616 TATATGGTGACAGA---GTCGATACAAGGACTTGGAAAGTAGTTGAAGACAAAGAAAT 3672
QY 185 -----IleAlaTrpAspLeu----- 189
Db 3673 GACAGATATTGTTTAAAGCTCATCAGAGCTCCGGAAGATTAAAGCTGGATCAACAA 3732
QY 190 -----GlnLysArgLysCysAspThrLeuLysSerLysLysLeuAsnIleArgValGly 206
Db 3733 GAATTTAATGAAGACGCAATGATTAACACCGTAACCTAGTTTAAACTTA----- 3783
QY 207 ArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluVal 226
Db 3784 -----TGCGGT----- 3789
QY 227 HisValGlyPheSerSerLysPheArgAspGluGlu-----GluSerPheThrLeuLeu 244
Db 3790 -----AAAAAATATTTCGATGAGCAATTACTAGAGAAAATACTTCTCT 3834
QY 245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
Db 3835 TCATATCAGGATGCATTAATAGTACTAAGGATTG----- 3870
QY 265 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle-----Phe 282
Db 3871 CTCAGCGCCCTCCAAAAGGTGAAGTACACAACTTTGCTGATATGGCGGTGGCACAATC 3930
QY 283 SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp 302
Db 3931 ACAACTTTCGACGGTTCGTTGTTGCTGATGTTCTTTAGCAATTAGCTATGAA--GAT 3987
QY 303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322
Db 3988 AAAAAGTGG---TCTTAGAAATCATCTTAGAATCAATATTTCATGACGCGAGAC----- 4038
QY 323 LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMet 342
Db 4039 -----AATATGAATGGGACAGCTGCACATGATAAGACTGATTAT----- 4077
QY 343 AlaSerHisGlyThrGlySerAlaAla----- 351
Db 4078 -----TATGGCTCTGAGCATCGCATTTGCTTACCTAAGTATTATTGATTAGATCTA 4131
QY 352 ---LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 370
Db 4132 GATTTCAGAACAAATAGAAATAT-----CTGAAGTTTGGGTTA----- 4167
QY 371 AsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerVal 390
Db 4168 -----GCTACAGCATTAACGCATGAGAACCTGAACGTT 4200
QY 391 Ser-----AsnLysProAlaIleIleLeu 398
Db 4201 AGTCTGTGCTGCTAAGGAGTAAGAGAGTTCCTTTGGTCTAGACAGCGCTGAATTAGCA 4260
QY 399 SerSerLeuValGlyAsnLeuValAspGlnSerLys----- 410
Db 4261 TCCCGTTGTATAGTGGATAGTTGATGATGCCAGATTACGAAGGGAAGATAGTAGGTA 4320
QY 411 -----GlnGlyIleValPheAsnGluAla-----SerTrpAlaGln 422
Db 4321 CGTAGGTTTTTATCATTTTACAGGTTGCTGAATTCAGACCCGCACTCGAAAATGGAATGAT 4380
QY 423 Leu-----ArgArgGluLeuLeuGlyGly-----AlaLeuSerLeuProAsp 436
```


Search completed: March 31, 2004, 13:13:05
Job time : 138.859 secs

Db 9558 GTTAAACAATTAAAGGAGCTTTTAAAT-----CATGTGTTATCCGAGATGGAATTG 9611
QY 281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
||||| : : : : :
Db 9612 ATATTGTAATCAGAGATCGATT-----AAAGAATTCTCGAGTGTAGATTAC--- 9659
QY 301 GlyAspMetAlaTrpValCysTrpAspProGluIleValaspGlyPheValAsnAlaGlu 320
||| : : : : :
Db 9660 -----ATTGATAAAATCTTAGATATATTTAATAAATCTACA 9695
QY 321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPhe----- 338
||| : : : : :
Db 9696 ATAGATGAATACAGAGATTCTCTTTTITAGA-----ACATTGGGCAT 9743
QY 339 LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys-----Glu 353
||| : : : : :
Db 9744 CCTCCATTAGAGCTAGTATTGCAGCAGAAAGTTAGAAATATATATATATGCGAAA 9803
QY 354 GlnThrThrTrpAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeu 373
||| : : : : :
Db 9804 CAATTAAATTTGACACTATTATAAATGTCATGCTATC----- 9842
QY 374 GlyMetCysThr-----AsnTyrLysGluArgLeuCys----- 384
||| : : : : :
Db 9843 ---TTCTGTACAATAATAATTAAACGGATATAGAGAAAGCAGTGGTGGACAGTGGCTCCT 9899
QY 385 -----TyrIleAsnSerValSerAsnLysProAla 395
||| : : : : :
Db 9900 GTGACATTACTGATCATGCACAGCAATTCATATAATGCTTACGGTTCAAATTCGCG 9959
QY 396 IleIleLeuSerSerLeuValGlyAsnLeuValAsp-----GlnSerLysGlnGlyIle 413
||| : : : : :
Db 9960 ATATCATATGAA-----AACGCTGTTGATTATTACCAGAGCTTTATAGGAATA 10007
QY 414 ValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyAlaLeuSer 433
||| : : : : :
Db 10008 AAATTTAATAAATTCATAGAACCTCAGTTCATGATGAAGATTG-----ACAATTTAT 10056
QY 434 LeuProAspProMetTyr-----LysSerAspSerTrp----- 444
||| : : : : :
Db 10059 ATGAAGATTAAGCATGTCTCCAAAATAATCAAACTGGACACAGTTTCTCTGCATCT 10118
QY 444 ----- 444
Db 10119 AATTACTGTACCGTACTAACGCATCCACGAATCACGAAGATTAGTTGAAAAATTATA 10178
QY 445 ---LeuGlyArgGlyGluProThrHisIleIleAspTyrLeu----- 457
Db 10179 GCAGATAGTAATTTGATCTCTAATCAGATATTAGATTAGTAGAATCTGGGACTGGTTA 10238
QY 458 -----LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHis 474
||| : : : : :
Db 10239 GATGATCCAGATTTAATAATTTCTTATAGTCTTAAAGAAAAAGAGATCAAA----- 10289
QY 475 AsnAlaMetLysAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu 494
||| : : : : :
Db 10290 -----CAAGAAGGT----- 10298
QY 495 AlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu 514
||| : : : : :
Db 10299 -----AGACTCTTTGCAAAAAATGACATATAAATGAGAGCTACACAGTTTAA 10346
QY 515 Phe---ThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal 533
||| : : : : :
Db 10347 TCAGAGACACTACTTGCATAATAATATAGGAAATCTTTCAAGAAAAATGGGATGGTAAAA 10406
QY 534 LysAsnLysGluMetArgAspSerLysAspProTyrProVal-----ArgVal 549
||| : : : : :
Db 10407 GGAGAGATTGAATTACTTAGAGATTAAACAACCATATCAATATCAGAGATTCCACGGTAT 10466
QY 550 AsnGlnValTyr 553
||| : : : : :
Db 10467 AATGAGATATAC 10478

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - nucleic search, using frame_plus_p2n model
Run on: March 31, 2004, 06:26:16 ; Search time 2230.2 Seconds
(without alignments)
7672.409 Million cell updates/sec

Title: US-09-913-878A-2_COPY_710_1282
Perfect score: 3010
Sequence: 1 GVGRMSRVAKIRDVILGL.....EKWCAITPEAMDKSGANYDS 573
Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Command line parameters: -DB=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USFT001/US9913878/runat_29032004_082008_8379/app.query.fasta_1.2254
-DB=EST -CFMP=fastap -SUFFIX=1st -MINMATCH=0.1 -LOCPCL=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US9913878 @CNC 1 1 4565 @runat_29032004_082008_8379 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc1:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc1:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|-------------|
| 1 | 240.5 | 8.0 | 2199 | 11 | AY103827 | Zea mays |
| 2 | 204 | 6.8 | 883 | 14 | CA765940 | AF53-Rpf |
| 3 | 191.5 | 6.4 | 577 | 10 | BF096404 | EST360431 |
| 4 | 184.5 | 6.1 | 954 | 29 | CG445102 | OG8A57TV |
| 5 | 183.5 | 6.1 | 841 | 14 | CA257764 | SCCCFL110 |
| 6 | 175.5 | 5.8 | 582 | 12 | B1246074 | IP1_68.D1 |
| 7 | 173 | 5.7 | 622 | 13 | B0965311 | eat08907 |
| 8 | 172 | 5.7 | 761 | 10 | BF275078 | GA_EB002 |
| 9 | 171 | 5.7 | 934 | 29 | CG212111 | OGOFJ51TV |
| 10 | 169.5 | 5.6 | 648 | 14 | CA223511 | SCJFFL1C0 |
| 11 | 169.5 | 5.6 | 741 | 14 | CF441078 | EST677423 |
| 12 | 168 | 5.6 | 541 | 13 | BQ827639 | Qd58f03.Y |
| 13 | 168 | 5.6 | 677 | 14 | CG928156 | GR45.104C |
| 14 | 167.5 | 5.6 | 916 | 29 | CG223171 | OGWALI4TV |
| 15 | 166 | 5.5 | 812 | 29 | CG134248 | WHERA369.F |
| 16 | 166 | 5.5 | 812 | 29 | CG140649 | FUIHY65TB |
| 17 | 164 | 5.4 | 570 | 10 | BE433646 | EUI400175 |
| 18 | 163.5 | 5.4 | 585 | 14 | CF244986 | 3530.1.8 |
| 19 | 163 | 5.4 | 640 | 10 | BF460205 | 069G02.Ma |
| 20 | 159.5 | 5.3 | 741 | 13 | BQ997497 | QGG16J23 |
| 21 | 157.5 | 5.2 | 621 | 13 | B0997627 | H108116r |
| 22 | 154.5 | 5.1 | 629 | 12 | B0268864 | B0268864 |
| 23 | 154 | 5.1 | 781 | 29 | CG938103 | MBENF75TR |
| 24 | 153.5 | 5.1 | 885 | 28 | BH163571 | ENTB82TF |
| 25 | 151 | 5.0 | 658 | 28 | BZ336439 | h234c12.b |
| 26 | 150 | 5.0 | 352 | 14 | CF053563 | OCN18d04 |
| 27 | 145 | 4.8 | 351 | 14 | CF053707 | OCN18d04 |
| 28 | 144.5 | 4.8 | 967 | 14 | CA272420 | SCVFLB208 |
| 29 | 143.5 | 4.8 | 411 | 10 | BF252757 | EST420020 |
| 30 | 143 | 4.8 | 342 | 14 | CF055747 | OCN3909.Y |
| 31 | 143 | 4.8 | 633 | 14 | CA199990 | SCRFLL103 |
| 32 | 140.5 | 4.7 | 1133 | 14 | CK210907 | FGA302273 |
| 33 | 140 | 4.7 | 442 | 28 | BH739133 | BOHNT59TF |
| 34 | 140 | 4.7 | 723 | 12 | BG594867 | EST49345 |
| 35 | 139 | 4.6 | 295 | 28 | BZ992509 | EST328265 |
| 36 | 139 | 4.6 | 613 | 14 | CA220413 | SCSBFL401 |
| 37 | 138.5 | 4.6 | 597 | 10 | AW649811 | AW649811 |
| 38 | 138.5 | 4.6 | 895 | 12 | BG344709 | HVSMEG001 |
| 39 | 136.5 | 4.5 | 525 | 10 | AW933692 | EST359535 |
| 40 | 136 | 4.5 | 405 | 10 | BF459767 | 064G04.Ma |
| 41 | 134.5 | 4.5 | 517 | 12 | BM368419 | EBB08.SQ |
| 42 | 131 | 4.4 | 554 | 14 | CD96715 | GL74.103L |
| 43 | 130.5 | 4.3 | 734 | 14 | CF449332 | EST685677 |
| 44 | 128.5 | 4.3 | 610 | 10 | BE920043 | EST423812 |
| 45 | 128 | 4.3 | 222 | 14 | CA902339 | PCSC13733 |

ALIGNMENTS

RESULT 1
AY103827
LOCUS AY103827 2199 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PC0135426 mRNA sequence.
ACCESSION AY103827
VERSION AY103827.1 GI:21206905
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2199)

AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2199)
AUTHORS Coe, B.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones,
 these are publicly available from ZmDB and may be found by BLAST
 searching at MBL, maizenap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES

source
 1..2199
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="MaizeDB:637927"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

ORIGIN

Alignment Scores:
 Pred. No.: 3,14e-17 Length: 2199
 Score: 240.50 Matches: 137
 Percent Similarity: 36.87% Conservative: 96
 Best Local Similarity: 21.68% Mismatches: 258
 Query Match: 7.99% Indels: 141
 DB: 11 Gaps: 23

US-09-913-878a-2_COPY_710_1282 (1-573) x AY103827 (1-2199)

Qy 1 GlyValGlyArgMetSerArgSerValAla-----Lys 11
 Db 171 GGAGTTGGAAAGATCTCAGCTAATTTTGCAGTGGAGGTGGCTATGAAGTGCAAAATTGAAA 230
 Qy 12 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 31
 Db 231 CGCTTT-----GCTCCTTCTGTTTTCAGATAAGGTATGCG 266
 Qy 32 SerAlaLysGlyMetTrpValIleAspValAspThrGlyAspGluAspTrpIleGlu 51
 Db 267 GGTTACAAAGTGTGTCGGTGTA-----GATACAGATCAAAATCAATGAAGCTTTCT 317
 Qy 52 ThrTyProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 71
 Db 318 TTGAGAAAAGCATGTCAAAGTTCAGTCCAGTCAAGAAATATC-----ACTCTT 362
 Qy 72 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuPro 91
 Db 363 GATGCTCTTCATAC---AGCAAGTACCAACCATGCTTCTGATCGGAGTTGATTACT 419
 Qy 92 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 111
 Db 420 CTTCTCTCA-----ACACTTGGGGTTAGCGATAATGTC 452
 Qy 112 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 131
 Db 453 TTTGAGTAAGCAGAGGAGGCTTAAGCCAG-----TTGACAGAGATGCTTAAT 503
 Qy 132 PheArgGlnTrpValTyGluSerTyProSerArgAlaThrArgValSerHisGlyArg 151

Db 504 -----GAACACACAGGCTGCTGCTGAAGCAGTGAACATT 536
 Qy 152 ValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsn 171
 Db 537 ATGCCCATG-----GAGAAAGTAACCAATGTAGTTAAAGAAATTGTGTCA 581
 Qy 172 SerGlyPheAspProLysLysGlnLysTyLeuGlnAspIleAlaTrpAspLeuGlnLys 191
 Db 582 TGTGGCTACACAGCCTGATCATGAGCCATATCTTCCATGCTGTCAAACTTTTAGAGCA 641
 Qy 192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyIle 211
 Db 642 TCCAAAGCTTCTAGAGTTGAACAAAGTCAAGGATTTTCATCACAGGGCCGAGCAATG 701
 Qy 212 TyrMetIleAlaAspPheTrpGlyValLeuGlnGluAsnGluValHisValGlyPheSer 231
 Db 702 ATGGGTTCCCTCGATGAACCTGCACACTTAAGTACGGCCAGGTATTTCTCCAAAGTTCT 761
 Qy 232 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 251
 Db 762 TACAGTCAGATCAGCATCGCAAG-----GTCGTTGTAACCTGGAAGAGTAGTTGTC 812
 Qy 252 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 271
 Db 813 GCCAAATATCCTTGTCTCCACCTGTGTGACATACGGGTTCTCCAGGCTGTGTATGTTCT 872
 Qy 272 GluLeuHisSerLysLysAspValIlePheSerThrLysGlyAspValProLeuAla 291
 Db 873 GCTCTGCACCACTGTTTGTACTGTGTCTTCCACAGCAGGAGCCAGGCCGCCACCT 932
 Qy 292 LysLysLeuSerGlyGlyAspTyArgAspGlyAspMetAlaTrpValCysTrpAspProGlu 311
 Db 933 AATGAGGTTTCAGGAGTGTATCTTGATGGGACATATATTTCTTCTGGGATCCACAT 992
 Qy 312 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyLeu 331
 Db 993 CTTATTCCAAAGTCGTTTGTGGATCCCTATGACATATCTCCAGCTTCAGCAGAAACATTA 1052
 Qy 332 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 351
 Db 1053 GACCATGAT----- 1061
 Qy 352 LysGluGlnThrThrTyArgMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371
 Db 1062 -----GTCACATATTGAGGAGATACAGGAGTACTTCAAACTACATAGTTAATGAG 1112
 Qy 372 PheLeuGlyMetCysThrAsnTy-----LysGluArgLeuCysTy 385
 Db 1113 AGCTTGGGATATCGCCATGCGCATGCTCTTTACACATCAGGACGATG----- 1166
 Qy 386 IleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeu 405
 Db 1167 -----AAAGCTGAGAGTCCACCGTGGCTTCAACTGGCCAAAGCTTCTCTATAGCT 1217
 Qy 406 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 425
 Db 1218 GTCGATTTCCCAAGACTGGAGTGGCG-----GCTCTGATTCACAT 1259
 Qy 426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet----- 438
 Db 1260 GAGCTA-----CATGTCAAGGAGTATCTGATTCATGAGGAAACTCGCAAAAGTCACC 1313
 Qy 439 TyLysSerAspSerTrpLeuGlyArg-----GlyGluProThrHisIle 453
 Db 1314 TATGAATCAAGGGTGTATCGGGAGCTCTATAGGAAATAAGAGAGCACACACAC 1373
 Qy 454 IleAspTyLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu-----Glu 471
 Db 1374 ATAAAGCACTTCACAGGGAAGTGGCAAGGGGTCTTATCACACCGATTTGATTGTAT 1433
 Qy 472 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 491
 Db 1434 GGCTATGAAGATTACATTACTGAGGCTATAGATTCAAGGAAGAGTAGCTTCAGGCTG 1493

QY 492 ProAspLeuAlaSerTyrThrPhePheYsGlu-----IleSerAspYsSer 508
 Db 1494 GGTAACTCTATGACCATATGGCTATAAAAGTGAAGTGAATTAAGTGGATGATT 1553
 QY 509 ArgSerSerAlaLeuLeuPheThr-----ThrLeuYsAsnArgTle 522
 Db 1554 CTAAGATGGCAAGAANAATTCACCAAGAGTAGTGCTGATGCAATTAGAATGGCGTG 1613
 QY 523 GlyGluValGluYsGlu-----TyrGlyArgLeuValYsAsnYsGluMet 538
 Db 1614 AGATCTTTGAGGAAGAAGCTAGTGTGCTTCAATGAGATGACGACAGGAGGATGGC 1673
 QY 539 ArgAspSerYsAspPro-----TyrProValArgValAsnGlnValTyr 553
 Db 1674 CAAGATGCCATGGAGCCAGGCCTCTGCTGGTACCATGTACTTATCATCAGCATAC 1733
 QY 554 -----GluYsTyrCysAla 558
 Db 1734 TGGGCAGCTACAATGAAGGTATGATCGCGGCATCTTATTAGCTTCCCATGGCGTA 1793
 QY 559 IleThrProGluAlaMetAspYsSerGlyAlaAsn 570
 Db 1794 TATGACAAAGCTGTGGCCATCAAGCAGGGAGGAAT 1829

CA765940 883 bp mRNA linear EST 08-JAN-2003
 AF53-Ref 09 H23 T7 092 ab1 IRRI Drought Stress Panicle Library
 Oryza sativa (indica cultivar-group) cDNA clone C0003263 5' similar
 to unknown, mRNA sequence.

CA765940
 CA765940.2 GI:27547887

RESULT 2

CA765940

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

On Dec 2, 2002 this sequence version replaced gi:25995195.
 Contact: Richard Bruskiewich
 Biometrics and Bioinformatics Unit
 International Rice Research Institute
 DAPO 7777, Metro Manila, Philippines
 Tel: +63-2-845-0563
 Fax: +63-2-845-0606
 Email: r.bruskiewich@cgiar.org
 International Rice Information System (IRIS)
 http://www.iris.irri.org: D0203262
 Assignment of putative function to the sequence by S. Rudd of the
 Munich Information Center for Protein Sequences
 (http://mips.gsf.de)
 Plate: 09 row: H column: 23.
 Location/Qualifiers
 1..883

FEATURES

source

/organism="Oryza sativa (indica cultivar-group)"

/mol_type="mRNA"

/cultivar="IR64"

/db_xref="taxon:39946"

/clone="C0003263"

/tissue_type="Panicles"

/dev_stage="Flowering"

/clone_lib="IRRI Drought Stress Panicle Library"

/note="Vector: pBluescript II SK+; Water stress was applied by not watering for 4 consecutive days. Panicles were collected from control (well watered) and stressed plants at 2 days before heading, at heading, and 50% flowering and 4 days after 50% flowering."

ORIGIN

Alignment Scores:
 Pred. No.: 1,96e-13 Length: 883
 Score: 204.00 Matches: 56
 Percent Similarity: 46.26% Conservative: 43
 Best Local Similarity: 26.17% Mismatches: 99
 Query Match: 6.78% Indels: 16
 DB: 14 Gaps: 3

US-09-913-878A-2_COPY_710_1282 (1-573) x CA765940 (1-883)

QY 115 LeuGlnArgGlnPheSerGluGlnYsHisAlaLeuAsnArgProValGluPheArgGln 134
 Db 45 TTAATCGACAGATTATAATATTGCTATCTCTCACTGAATGTCCTCCAGATTCTATCTTTGG 104
 QY 135 TrpValTyrGluSerTyrSerArgAlaThrArgValSerHisGlyArgValProphe 154
 Db 105 CAAATGACAGAGACCATGCTTTCAACCTCAACAATATTCTATCATCAGACAGATGTTGCT 164
 QY 155 LeuAlaGlyLeuProAspSer-----GlnGluGluThrLeuAsnPheLeuMetAsn 171
 Db 165 TTTGAGGTTTAAACAATCTCATGCTGCTGATGGAACACTGCAGCATTTGCTCAGT 224
 QY 172 SerGlyPheAspProYsLysGlnYsTyrLeuGlnAspIleAlaTrpAspLeuGlnYs 191
 Db 225 GCTGGCTTTGAACCTAGAACTGAACCACTTTGAAAGCAATGCTCTTGGCTATAGGTCT 284
 QY 192 ArgGlyCysAspPheLeuYsSerYsLeuAsnIleArgValGlyArgSerAlaTyrIle 211
 Db 285 GCACAATTGCGAGATCTTTTGGNAAAAGCAAGATATTTGTGCCAAGGAGGTGGTTC 344
 QY 212 TyrMetIleAlaAspPheTrpGlyValLeuGluGlu----- 223
 Db 345 ATGGGCTGCTTTCATGAGCTTGGGTTCTTGGACAAAGCGAGTCTTTATTTCGGGCAACA 404
 QY 224 -----AsnGluValHisValGlyPheSerSerYsPheArgAspGluGluGlu 239
 Db 405 GTTCCATCATTTGAATAGTATTTTTGAAGCATGGGTCAAGATTTTCATCAACAGATAAA 464
 QY 240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
 Db 465 AACACAGAGGTCTATTTTGGGA---ACTGTGTGTAATAGCAAGAATCCCTGTCTTCATCCA 521
 QY 260 SerAspIleGlnArgValArgAlaValPheYsProGluLeuHisSerLeuYsAspVal 279
 Db 522 GCGGATGTCGGCATCTCTGAAGCAGTTGATGTCGCCAACTGCATCATCTGTTGTTGTTGT 581
 QY 280 IleIlePheSerThrYsGlyAspValProLeuAlaYsLysLeuSerGlyGlyAspTyr 299
 Db 582 TTGGTGTTCCTCCAGAAAGGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTT 641
 QY 300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
 Db 642 GATGGGGATCTCTACTTTGTGACATGGGATGAGAACTTATA 683

RESULT 3

BF096404

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BF096404 577 bp mRNA linear EST 18-MAY-2001
 EST360431 tomato nutrient deficient roots Lycopersicon esculentum
 cDNA clone cLEW106 5' sequence similar to RNA-directed RNA
 polymerase {Arabidopsis thaliana}, mRNA sequence.

BF096404

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

LOCUS

DEFINITION

| Db | 482 ATA 484 |
|--|---|
| Unpublished (2001) Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@cuga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTMix or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV High quality sequence stop: 560 POLYA=No. | BU965311 622 bp mRNA linear EST 21-OCT-2002 sat08g07.v1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-cl036-13357 5' similar to TR:Q9ZRS8 Q9ZRS8 RNA-DIRECTED RNA POLIMERASE ;, mRNA sequence. BU965311 BU965311.1 GI:24206058 EST. BU965311 SOURCE Glycine max (soybean) ORGANISM Glycine max |
| RESULT 7 LOCUS DEFINITION | |
| ACCESSION VERSION KEYWORDS SOURCE ORGANISM | |
| REFERENCE AUTHORS | 1 (bases 1 to 622) Shoemaker,R., Keim,P., Vodkin,L., Erpelidng,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Pearson,B., Swaller,T., Gibbons,M., Pape,B., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,X., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. Public Soybean EST Project Unpublished (1999) Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: cu@resgen.com web site: www.resgen.com Seq primer: -40RP from Gibco High quality sequence stop: 445. |
| TITLE JOURNAL COMMENT | |
| FEATURES Source | 1..622 /organism="Glycine max" /mol_type="mRNA" /db_xref="taxon:3847" /clone="SOYBEAN CLONE ID: Gm-cl036-13357" /tissue_type="somatic embryos cultured on MSD 20" /lab_host="DH10B" /clone_lib="Gm-cl036" /note="vector: pSPORT1; Site_1: NotI; Site_2: Sall; This cDNA library was constructed from mRNA isolated from somatic embryos (age ranging from 2 months to 9 months) cultured on MSD 20. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restriction site. Sall linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-Sall restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. e-mail: l-vodkin@uiuc.edu" |
| Alignment Scores: Pred. No.: Score: Percent Similarity: Best local similarity: Query Match: DB: | 6.67e-10 Length: 622 173.00 Matches: 44 50.00% Conservative: 34 28.21% Mismatches: 70 5.75% Indels: 8 13 Gaps: 2 |
| ORIGIN | |
| 169 LeuMetAsnSerGlyPheAspProLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 188 17 ATGTTGAGTGGCTGATTCACCTGGACTGACACACCTGAAGCAATGCTTTAGCT 76 189 LeuGlnLysArgGlyCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 208 177 ATAAGGTCCTCAGACTACTGGTCTTTTGGAGAGACAGAGATTTTGTGCCAGAGGGG 136 209 AlaTyrIleTyrMetIleAlaAspPheThrGlyValLeuGluGlnAsnGluValHisVal 228 137 AGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 186 229 -----GlyPheSerSer 232 197 CGGGCGTCATCTCCATCACTCAATATGTCGTGTAAGCATGACCAAGATTTTCTCTCA 256 233 LysPheArgAspGluGluSerPheThrLeuLeuSerAspCysAspValLeuValala 252 257 GCACACAAAGTGCAGAG-----ACCATGTGGGT-----ACTATCGTAAATGGCA 301 253 ArgSerProAlaHisPheProSerAspIleGlnArgValaGalaValPheLysProGlu 272 302 AAGAATCCATGCTTCATCCAGGATGTCGGATCCTTGAAGCTGTGTGATGTCCTGAA 361 273 LeuHisSerLeuLysAspValIlelePheSerThrLysGlyAspValProLeuValLys 292 362 CTGCATCACCTTTGTTGATGCTTGTGCTTCTCCCAAGAAAGGTGAGAGCCACATGCCAAT 421 293 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTyrAspProGluIle 312 422 GAAGCATCTGGAGATGATCTTGGATCTTCTACTTCTTGTGACATGGATGAAACCTT 481 313 Val 313 ::: | |

An unidirectional cDNA library generated from [inflorescence at beginning of development (1cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucess.lad.ic.unicamp.br/public>

ORIGIN

Alignment Scores:
 Pred. No.: 1,91e-09 Length: 648
 Score: 169.50 Matches: 59
 Percent Similarity: 41.91% Conservative: 42
 Best Local Similarity: 24.48% Mismatches: 111
 Query Match: 5.63% Indels: 29
 DB: 14 Gaps: 6

US-09-913-878A-2_COPY_710_1282 (1-573) x CA223511 (1-648)

QY 85 LeuAenLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArg 104
 DB 11 CTGAATCGAAGTGAATCTCTCTCA-----ACA 43
 QY 105 GlnAlaIleGlyAspArgLeuLeuAsnAspLeuGlnArgGlnPheSerGluGlnLysHis 124
 DB 44 CTGGGGTGGCAGTAGTCTTCACTAAGCAGAGGAGCTGAAGCAG----- 97
 QY 125 AlaLeuAenArgProValGluPheArgGlnTrpValTyrGluSerTyrSerArgAla 144
 DB 98 ---TTGAACAGAAATGTAACC-----GAACACACAGGCT 127
 QY 145 ThrArgValSerHisGlyArgValPheLeuAlaGlyLeuProAspSerGlnGluGlu 164
 DB 128 GCTCTGAAGCGGTGAACCTATGCCCATG-----GGAGAAGTTACCAAT 172
 QY 165 ThrLeuAenPheLeuMetAsnSerGlyPheAspProLysGlnLysTyrLeuGlnAsp 184
 DB 173 GPAGTTAAGAAGTGTATCATGTGCTACCGACCTGATCATGAGCGGTATCTTCCATG 232
 QY 185 IleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArg 204
 DB 233 CTGTACAAACTTTTAGAGCATCCAGCTTCTAGAAATTGAACAAAGTCAGGATATTC 292
 QY 205 ValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsn 224
 DB 293 ATCCACAGCGCGAGCATGATGCTGCTGGATGAACCCGACACTAAGTATGCG 352
 QY 225 GluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeu 244
 DB 353 CAGGTATTATCCAA---GCTTCTTACTGTGCAGATGACCATCGCAAGTTCGTTGAATC 409
 QY 245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
 DB 410 GGA-----AAGTAGTTGTTCCTCAAAATCTTCTCTCCACCTTGGTGCATACGGGTT 463
 QY 265 ValArgAlaValPheLeuProGluLeuHisSerLeuLysAspValIleIlePheSerThr 284
 DB 464 CTCAAGCTGTTGATATTCCTGCTCTGCACCACTTGTGTGATCTGCTGTTCTTCCACAG 523
 QY 285 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 304
 DB 524 CAGGACCAAGCCGACCCCTAATCAGTGTTCAGGAGTGAATCTTGTATGGGACATATAT 583
 QY 305 TrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu 324
 DB 584 TTGTGTTCTGGATTACATCTTATCCAAAGTCGTTTGGTGGATCTCTATGGACTATTCT 643
 QY 325 Pro 325
 DB 644 CCA 646

RESULT 11

CF441078
 LOCUS
 DEFINITION
 EST677423 normalized cDNA library of onion Allium cepa cDNA clone
 ACAGB56, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Allium cepa (onion)
 Allium cepa
 Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Allium.
 REFERENCE
 1 (bases 1 to 741)
 Haver, M.J., Cheung, F., Van Aken, S., Utterback, T. and Town, C.D.
 AUTHORS
 TITLE
 Expressed Sequence Tags from a normalized library of mixed onion
 tissues (Allium cepa)
 JOURNAL
 COMMENT
 Unpublished (2003)
 Contact: Haver MJ
 Department of Horticulture
 USDA-ARS and University of Wisconsin
 1575 Linden Drive, Madison, WI 53706, USA
 Tel: 608-262-1830
 Fax: 608-262-4743
 Email: mjhaver@facstaff.wisc.edu
 TIGR sequence name ACAGB56TK. For more information:
<http://haverlab hort.wisc.edu>
 Seq primer: CAG GAA ACA GCT ATG ACC.

FEATURES

source
 1..741
 /organism="Allium cepa"
 /mol_type="mRNA"
 /cultivar="Red Creole (bulbs), unknown (callus), Ebano &
 Texas Legend (roots)"
 /db_xref="taxon:4679"
 /clone="ACAGB56"
 /tissue_type="Callus, roots, and young bulbs"
 /clone_lib="normalized cDNA library of onion"
 /note="Vector: pCMVSPORT6.1-cdb (Invitrogen); Site.1:
 EcoRV (5'); Site.2: NotI (3'); Equal molar amounts of mRNA
 from callus, roots, and young bulbs were combined to
 synthesize the library. Normalization to enrich for
 low-copy transcripts was performed by proprietary
 techniques of Invitrogen."

ORIGIN

Alignment Scores:
 Pred. No.: 2,39e-09 Length: 741
 Score: 169.50 Matches: 64
 Percent Similarity: 40.94% Conservative: 49
 Best Local Similarity: 23.19% Mismatches: 106
 Query Match: 5.63% Indels: 57
 DB: 14 Gaps: 8

US-09-913-878A-2_COPY_710_1282 (1-573) x CF441078 (1-741)

QY 197 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 216
 DB 3 CTGCGCATAAACACCCAGGATATTGTGAATAAAGGAGGTGTGATGGATGTTTGGAT 62
 QY 217 PheTrpGlyValLeuGluAenGluValHisValGlyPheSerSer----- 232
 DB 63 GAACCTGGGCTGTAGAGCAAGGGCAGTGTCTTATTCAGGTTTCCAGTCTCTGTGTGAG 122
 QY 233 -----LysPheArgAspGluGluSerPheThrLeuLeu 244
 DB 123 AATGCCCTTTTGAAGTATGGGAGTCTTTTAAAGATAGGAAGAGAGAGAAAGAGGTGATA 182
 QY 245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
 DB 183 AAGGCT---ACAATAGTTTTCGCCAAGAATCCATGTTGCACCTCGGTGATGTAGGGTA 239
 QY 265 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 284

```

Db 240 CTTGAGCGGATGATGTCGCGAATTGAAGCATCTGTGGAGTGTGTTGTTTCTTCAG 299
Qy 285 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 304
Db 300 AAAGAGAGATAGACCCGATTAATAGGCTAGTGGACGACCTTGTATGGGATTTGTAT 359
Qy 305 TrpValCysTyrAspProGluLeuValAspGlyPheValAsnAlaGluMetProLeuGlu 324
Db 360 TTTGTTACTTGGATGAGAGGTTGATTCACCGGGTAAAGCAGTTGGCTCCGATGGAT 419
Qy 325 ProAspLeuSerArgTyrLeuLysLysAspLysThrPheLysGlnLeuMetAlaSer 344
Db 419 ----- 419
Qy 345 HisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 364
Db 420 ---TATACTGCAGCTGAGCGGAGGATGAGCCA-----CAACCCATTGTT 461
Qy 365 HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCys 384
Db 462 CATCAAGATATTAACTTCTTCATCAAGAACGTATCGAGT-----GAGAACCTAGGA 515
Qy 385 TyrIleAsnAsnSer-----ValSerAsnLys 393
Db 516 AGCATCAGCATGCTCATGATCTCATGTCAGACCAAGCATGATGGTCCATGACGAA 575
Qy 394 ProAlaIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly--- 412
Db 576 AATATGATCAAACTTGCAGAGCTTGGCGCTGAGCAGTGCAGCTCCCTTAAACCGGTAAA 635
Qy 413 ---IleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuGlyGlyAla 431
Db 636 ATAGTAAAGATGCTCGTCCCTCGTGAAGCCCAACTGATCCCTGACTTCATGGGCAAGGAC 695
Qy 432 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArg 447
Db 696 AGTTCGCTT-----ACCTACAAATCAGAGAAATATCTGGGCAGA 734

RESULT 12
BQ827639
LOCUS
DEFINITION
BQ827639.1 Moss EST library PPAS Physcomitrella patens cDNA clone
PEP SOURCE ID:PPAS041205 5' similar to TR:Q9SG02 Q9SG02 PUTATIVE
RNA-DIRECTED RNA POLYMERASE. ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Quatrano, R., Bashardes, S., Cove, D., Cuming, A., Knight, C.,
Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T.,
Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, F., Steptoe, M., Gibbons, M., Harvey, N., Ritter, S.,
Jackson, Y., McCann, R., Waterston, R. and Wilson, R.
Leeds/Wash U Moss EST Project
Unpublished (1999)
Contact: Ralph Quatrano
Leeds/Wash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Libraries were constructed by Dr. Stavros Bashardes as part of the
Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
Seq primer: -40RP from Gibco

```

High quality sequence stop: 447.
Location/Qualifiers

FEATURES

1..541

/organism="Physcomitrella patens"

/mol_type="mRNA"

/db_xref="taxon:3218"

/clone="PEP_SOURCE_ID:PPAS041205"

/dev_stage="protonemata, 7day old ABA-treated"

/lab_host="E.coli DH10b"

/clone_lib="Moss EST library PPAS"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: EcoRI; 7-day-old protonemal tissue was incubated 10-4M ABA for 16 hours before RNA isolation. cDNA was synthesised with from Amersham's cDNA Synthesis Plus kit and ligated with EcoRI-NciI linkers for cloning in the EcoRI site of Lambda ZapII (Stratagene). After packaging, the library was propagated in E. coli XL-I Blue cells and amplified. The library was excised by mass excision in SOLR cells and ampicillin resistant transformants selected. Approximately, 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered and used to transform DH10b cells by electroporation. Clones corresponding to abundant transcripts were identified by colony hybridization using a cDNA probe derived from untreated protonemal tissue, and eliminated from the library, by rearraying."

ORIGIN

Alignment Scores:
Pred. No.: 2.15e-09 Length: 541
Score: 168.00 Matches: 42
Percent Similarity: 44.79% Conservative: 31
Best Local Similarity: 25.77% Mismatches: 64
Query Match: 5.58% Indels: 26
DB: 13 Gaps: 3

US-09-913-878a-2_COPY_710_1282 (1-573) x BQ827639 (1-541)

Qy 169 LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 188
Db 68 ATGTTCCGCGAGGATTTCTCCCTTTGAGGAGCCCTCATCTCTATGAATTTACTGGTCTCT 127
Qy 189 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 208
Db 128 ATCAAGTTACGCTCGTGGAGGATTAGTTTCAAGCGCTAGATATTTGTACCTAATGGA 187
Qy 209 AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluAsnGlu-----Val 226
Db 188 CGTTGGGTTCATGGGTGTATGACGAGACTGGGTGTTTGAATTACGGGCAATGTTTCATT 247
Qy 227 HisVal-----GlyPhe 230
Db 248 CATGTATCTGGCCCTGTAAAGAGGCTGCTTGGAGGAGATGGAATAATCTGGGACAT 307
Qy 231 SerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeu 250
Db 308 AGTGACTCTCTGCAG-----GTGATCACGGGAAAGGTTATT 343
Qy 251 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 270
Db 344 ATGTTAAAAATCCATCTCTGAGATATCCGATATCCGATATCGGAGCGAGTGGATCTG 403
Qy 271 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 290
Db 404 CCGGAATTCACACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
Qy 291 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 310
Db 464 CCAATAGAGTCTCTGCGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 523
Qy 311 GluIleVal 313
Db 524 AATTGTATT 532

```

RESULT 13
CD928156
LOCUS
DEFINITION
ACCESSION
CD928156
VERSION
CD928156.1 GI:32775920
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum (bread wheat)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 677)
Genoplante, a major partnership french program in plant genomics
Unpublished (2003)
JOURNAL
COMMENT,
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
Location/Qualifiers
1..677
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="GR45104C10"
/tissue_type="grain (45 degrees per day after
pollination)"
/clone_lib="GR45"

FEATURES
source
1..677
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="GR45104C10"
/tissue_type="grain (45 degrees per day after
pollination)"
/clone_lib="GR45"

ORIGIN
Alignment Scores:
Pred. No.: 3,13e-09 Length: 677
Score: 168.00 Matches: 39
Percent Similarity: 49.71% Conservative: 47
Best Local Similarity: 22.54% Mismatches: 73
Query Match: 5.58% Indels: 14
DB: 14 Gaps: 2

US-09-913-878a-2_COPY_710_1282 (1-573) x CD928156 (1-677)
QY 155 LeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPhelLeuMetAsnSerGlyPhe 174
|||
|||
|||
45 CTGGGAAAATTGGTAGTCGAGAAACAAAGACAGCATCGAAAATTTTACTGCAAGGCTAT 104
QY 175 AspProLysGlnLysTyrLeuGlnAspIleAlaTyrAspLeuGlnLysArgLysCys 194
|||
|||
|||
105 GAACCAAGTTAGTGAGCTTACCTGTTGATGATTTTAAAGGCCCATCAGATTAATAGCTG 164
QY 195 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetile 214
|||
|||
|||
165 ACTGACATAAGAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 224
QY 215 AlaAspPheTyrGlyValLeuGluLysValHisValHisValHisValHisValHis 234
|||
|||
|||
225 TTGGATGAACAGAGTGAATAGATATGATGATGATGATGATGATGATGATGATGATGAT 284
QY 235 ArgAspGluGluGluSerPhe-----Thr 242
|||
|||
|||
285 AAAGAGCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 344
QY 243 LeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIle 262
|||
|||
|||
345 GCAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 404
QY 263 GlnArgValAlaValPheLysProGluLeuHis-----SerLeuLysAspValile 280
|||
|||
|||

```

```

Db 405 AGAGTACTTGAAGCTGTCTATGACCATGGTTTGTACGCTAAGAACCTGGTTGATTGT 464
QY 281 IllePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
Db 465 GTCTTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 524
QY 301 GlyAspMetAlaTyrValCysTyrAspProGluLeuVal 313
Db 525 GGTGACCTCTATTATTATCACTTGGGATGAGAAACTGATT 563

RESULT 14
CD923171
LOCUS
DEFINITION
OGWAL14TV ZM 0.7-1.5 KB Zea mays genomic clone ZMMBMA0519D04,
genomic survey sequence.
ACCESSION
CD923171
VERSION
CD923171.1 GI:34123059
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 916)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Frazer,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Other GSSs: CGWAL14TH
Unpublished (2002)
Consortium for Maize Genomics
Contact: Cathy Whitelaw
TIGR
912 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@cigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..916
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0519D04"
/clone_lib="ZM 0.7-1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

FEATURES
source
1..916
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0519D04"
/clone_lib="ZM 0.7-1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 5.98e-09 Length: 916
Score: 167.50 Matches: 59
Percent Similarity: 39.85% Conservative: 47
Best Local Similarity: 22.18% Mismatches: 101
Query Match: 5.56% Indels: 59
DB: 29 Gaps: 8

US-09-913-878a-2_COPY_710_1282 (1-573) x CD923171 (1-916)
QY 66 LysHisGlnArgThrLeuGluValArgSerValAlaSer-----GluLeuLysSerAla 83
Db 18 AAGTTCAGTCCAGCACACCGTGTCTGAAAGTGTCTGACCAAGCGTTCGCCAGCGCTTC 77
QY 84 GlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMet 103
Db 78 TTCTTGAACCCAGATCATACGCTCTCACCACCAAGCGTTCGCCAGCGCTTC 137
QY 104 ArgGlnAlaLysGlyArgArgLeuLeuAsnAspLeuGlnArgGlnPheSerGluGlnLys 123
Db 138 TGGCAG---CTGCAGGAAGCGATCTAGGCAAGCTCAGCGCGGTTTGTTCGACAGCAT 194

```



```

QY 124 HisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArg 143
Db 195 GTGGGG-----TACGAGTCTGACCAACTCT 221
QY 144 AlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGlu 163
Db 222 TGC-----CCTGAGCAAGGA 236
QY 164 GluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLysTyrLeuGln 183
Db 237 AGACCGCAGGCTGATGCTGGCGCTGCTAC----- 269
QY 184 AspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIle 203
Db 270 -----TGCCATCTCCACTCGCAGATCAGGCGCTTCTCAACAAATGTGATCT 320
QY 204 ArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGlu 223
Db 321 TTGTGCCCAAGGAAGGTGGCTGATGGGCTGTCTCGACAGTTTGGATCCCTTGAGCAG 380
QY 224 AsnGluVal-----HisValGly----- 229
Db 381 GGCAGTGTCTTCTCCGGTCTCGAGCCCATCAGTCGCGCAGCGCTGGTGAGAACGGTCT 440
QY 230 -----PheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCys 247
Db 441 CGGATATTCTCCCTCGAGTATAAGCGCGATGCGGATGCACAGTGATTACTTGGT----- 494
QY 248 AspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAla 267
Db 495 ACGTTGTGATGCGGAGAACCCATGCTTCATCCAGGGAGCGTCCGATCTCTTGAAGCT 554
QY 268 ValPheLysProGluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAsp 287
Db 555 GTGGATGTCCTGATCATCATTTGTCGACTGCTGCTCTCCCAAGAAAGGTGAG 614
QY 288 ValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCys 307
Db 615 AGGCCACAGCCCAAAAGCTTCCGGAGCGACTCGAGCGGCGAGTCTACTTCTGTGAG 674
QY 308 TrpAspProGluIleVal 313
Db 675 TGGGACGAGAGCCTTGTA 692

RESULT 15
LOCUS CF134248
DEFINITION WHE4369_F09_L17ZT Wheat meiotic floret cDNA library Triticum aestivum cDNA clone WHE4369_F09_L17, mRNA sequence.
ACCESSION CF134248
VERSION CF134248.1 GI:33219166
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 722)
Anderson, O.D., Chao, S., Crossman, C., Langridge, P., Lazo, G.R.,
Pham, J., Rauch, C. J., Sutton, T., Woo, J. and Wilson, C.
The structure and function of the expressed portion of the wheat
Genomes - Meiotic floret cDNA library
Unpublished (2003)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oanderson@wv.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: T7 primer.

```

FEATURES

```

1..722
Location/Qualifiers
1..722
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4369_F09_L17"
/tissue_type="Whole florets with anthers"
/dev stage="Meiotic stages pre-meiosis-metaphase I"
/lab_host="E. coli DH10B"
/clone_lib="Wheat meiotic floret cDNA library"
/notes="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Plants
were grown in a glasshouse. Anther meiotic stage was
determined microscopically after removing a single anther
from a primary floret. If determined to be between (and
including) meiotic stages pre-meiosis and metaphase I,
remaining floret tissue was collected and pooled for
library construction. The tissue, total RNA, and poly(A)
RNA were prepared, cDNA synthesised, and directionally
ligated into pSPORT1 by Tim Sutton in the P Langridge Lab
at the Department of Plant Science, University of
Adelaide, Waite Campus, Australia. Average insert size
1.4kb. Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."
```

ORIGIN

```

Alignment Scores:
Pred. No.: 6,12e-09 Length: 722
Score: 166.00 Matches: 51
Percent Similarity: 45.54% Conservative: 46
Best Local Similarity: 23.94% Mismatches: 102
Query Match: 5.51% Indels: 14
DB: 14 Gaps: 4

US-09-913-878a-2_COPY_710_1282 (1-573) x CF134248 (1-722)

QY 115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgPro---ValGluPheArg 133
Db 42 TTAACCGTCAGATCATACATTACTCTCTCTCTGTAATGTCCTGATCTATCTTTCA 101
QY 134 GlnTrpValTyrGluSerTyrSerArgAlaThrArgValSerHisGlyArgValPro 153
Db 102 CAAATGCAGAAAGACATGCTGTCTAATCTCAACATATTTTGTGACTGACTGATGTGCT 151
QY 154 PheLeuAlaGlyLeuProAspSerGlnGluGlu-----ThrLeuAsnPheLeuMetAsn 171
Db 162 TTTGACGTTGTACCACTCTTGTGTGACGAGGAAACACCCGACGACTGATGTGAGT 221
QY 172 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 191
Db 222 GCTGGCATTTTCACTGGAACTGAGCCGACCTGAAAGCATTTGCTGTGGTGTATGAGTCC 281
QY 192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 211
Db 282 TCACAGCTACTGGGCTCTTTTGGAAAATCAAGGATTTTGTTCCTCCAGGGAAGGTGTTG 341
QY 212 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 231
Db 342 ATGGCGCTCCCTTGATGAACCTTGGGATCTTGGACAGGACAGCATGCTTTATTCCGGGCTCA 401
QY 232 SerLys-----PheArgAspGluGluGluSerPheThrLeuLeuSer 245
Db 402 TCTCCAGTACTCAATAATTCTCTGCTGAGCATGGACCAAGATCTTCTTCAGCGAACAC 461
QY 246 AspCysAsp-----ValLeuValAlaArgSerProAlaHisPheProSer 260
Db 462 AATGCAGAGACCGTTAAAGGTACTCTTGTATGCAAGAATCCGTCCTCCCTTCATCCAGGG 521
QY 261 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
Db 522 GATGTCGGATCCCTTGAGGCCATTGATGTCCTGCTTGGATCACCTTGTGATGTGCTTG 581
QY 281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
```



```
Db      582  GTCTTCCCAAGATGGCGAGGCGCCACACGCCCAATGAGCGTCTGGGAGTGATCTTGAT 641
Qy      301  GlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
Db      642  GGGGATCTATACTTTGTGACGTGGGATGAAAACTTATC 680
```

Search completed: March 31, 2004, 13:00:01
Job time : 2240.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 15:33:39 ; Search time 19844 Seconds
(without alignments)
17571.795 Million cell updates/sec

Title: US-09-913-878A-1
Perfect score: 8045
Sequence: 1 ggtagcgggcccccctcga.....cgtattacaatccagtcac 8045

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:
2: gb_hgt:
3: gb_in:
4: gb_cm:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_ov:
22: em_or:
23: em_pat:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sts:
28: em_un:
29: em_vi:
30: em_hgt_hum:
31: em_hgt_inv:
32: em_hgt_other:
33: em_hgt_mus:
34: em_hgt_pln:
35: em_hgt_rod:
36: em_hgt_mam:
37: em_hgt_vrt:
38: em_sy:
39: em_hgt_hum:
40: em_hgt_mus:
41: em_hgt_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|------------|--------------------|
| 1 | 8045 | 100.0 | 8045 | 6 | AX034425 | Sequence |
| 2 | 4206 | 52.3 | 4206 | 8 | NCR13328 | Neurospora |
| 3 | 284.2 | 3.5 | 33321 | 3 | AC116986 | Dictyoste |
| 4 | 276.6 | 3.4 | 25215 | 2 | AC095632 | Rattus no |
| 5 | 273.4 | 3.4 | 229259 | 2 | AC096212 | Rattus no |
| 6 | 273.2 | 3.4 | 200412 | 2 | AC115666 | Rattus no |
| 7 | 273.2 | 3.4 | 222913 | 2 | AC120325 | Rattus no |
| 8 | 271.4 | 3.4 | 224733 | 3 | AC117075 | Dictyoste |
| 9 | 271 | 3.4 | 118689 | 2 | AC141123 | Rattus no |
| 10 | 271 | 3.4 | 205454 | 10 | AC096792 | Rattus no |
| 11 | 269.2 | 3.3 | 198573 | 2 | AC110118 | Rattus no |
| 12 | 268.6 | 3.3 | 255678 | 2 | AC099432 | Rattus no |
| 13 | 268.4 | 3.3 | 279242 | 2 | AC114079 | Rattus no |
| 14 | 268.4 | 3.3 | 346208 | 2 | AC128290 | Rattus no |
| 15 | 267.8 | 3.3 | 211295 | 2 | AC130618 | Rattus no |
| 16 | 267.2 | 3.3 | 110000 | 3 | AC116984_0 | Dictyoste |
| 17 | 267.2 | 3.3 | 110000 | 3 | AC116984_1 | Continuation (2 of |
| 18 | 266.6 | 3.3 | 240214 | 2 | AC126871 | Rattus no |
| 19 | 266.4 | 3.3 | 175351 | 2 | AC117900 | Rattus no |
| 20 | 266.4 | 3.3 | 247341 | 2 | AC091178 | Rattus no |
| 21 | 264.2 | 3.3 | 136240 | 3 | AC117070 | Dictyoste |
| 22 | 263.4 | 3.3 | 7921 | 3 | AF153362 | Dictyoste |
| 23 | 261.6 | 3.3 | 110000 | 2 | AC129796 | Rattus no |
| 24 | 261.6 | 3.3 | 240800 | 2 | AC095964 | Rattus no |
| 25 | 261.6 | 3.3 | 274697 | 2 | AC098240 | Rattus no |
| 26 | 261.2 | 3.2 | 282355 | 2 | AC104401 | Rattus no |
| 27 | 260.6 | 3.2 | 256528 | 2 | AC108270 | Rattus no |
| 28 | 260.4 | 3.2 | 177054 | 2 | AC126897 | Rattus no |
| 29 | 260.4 | 3.2 | 232357 | 2 | AC127133 | Rattus no |
| 30 | 260.2 | 3.2 | 136662 | 2 | AC129737 | Rattus no |
| 31 | 260.2 | 3.2 | 219107 | 2 | AC115259 | Rattus no |
| 32 | 260 | 3.2 | 96573 | 2 | AC136672 | Rattus no |
| 33 | 260 | 3.2 | 237900 | 2 | AC125525 | Rattus no |
| 34 | 260 | 3.2 | 245524 | 2 | AC102954 | Rattus no |
| 35 | 259.2 | 3.2 | 301081 | 2 | AC108658 | Rattus no |
| 36 | 259.2 | 3.2 | 347444 | 2 | AC111888 | Rattus no |
| 37 | 259 | 3.2 | 233094 | 2 | AC113774 | Rattus no |
| 38 | 258.2 | 3.2 | 217746 | 2 | AC112302 | Rattus no |
| 39 | 257.8 | 3.2 | 104708 | 2 | AC139336 | Rattus no |
| 40 | 257.8 | 3.2 | 241531 | 2 | AC129959 | Rattus no |
| 41 | 257.8 | 3.2 | 241765 | 2 | AC132557 | Rattus no |
| 42 | 257.8 | 3.2 | 256879 | 3 | AC116982 | Dictyoste |
| 43 | 257.8 | 3.2 | 278876 | 2 | AC120936 | Rattus no |
| 44 | 257 | 3.2 | 232373 | 2 | AC134805 | Rattus no |
| 45 | 256.8 | 3.2 | 284726 | 2 | AC096844 | Rattus no |

ALIGNMENTS

RESULT 1
AX034425
LOCUS AX034425 8045 bp DNA linear PAT 22-SEP-2000
DEFINITION Sequence 1 from Patent WO0050581.
ACCESSION AX034425
VERSION AX034425.1 GI:10303130
KEYWORDS Neurospora crassa
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1
AUTHORS Cogoni, C. and Macino, G.
TITLE Isolation and characterization of a n. Crassa silencing gene and
uses thereof

JOURNAL Patent: WO 0050581-A 1 31-AUG-2000;
 UNIV ROMA (IT) / COCONI CARLO (IT) / MACINO GIUSEPPE (IT)
 FEATURES Location/Qualifiers
 source 1..8045
 /organism="Neurospora crassa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5141"
 2447..6655
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC10121.1"
 /db_xref="GI:10303131"
 /db_xref="REMBL:CAC10121"
 /translation="MNPETPKRSPEVEINLRNLDYNIQLQCVADTILTPHRRKEL
 AESDDEFGRHDKIYRALNPLVYRKDDSLNQAEANFFLEAKAASNWVPKAAADPTDLP
 WSKGPPRAATQAGQAWQLVLELVNPPPPNPTPGRTLTLSGSPSLNLTSTNT
 KRQDEPNATVPADPQKSLTSATGPIIHGAALPKPDPVNTGSKPSLSESLNQC
 TKRAKGLDVAANAAPPVPIASGLQPKPRRHRANTDPTATGHRARDDQDFDTSQ
 GTSGVSSVFSACRNQSTQSSFPAPSQSPKEPVDATVEAGHILIESFKGRITKS
 HIDNQPLSSSQGISTSTYYESPSSGGEGALPEPSRNGLARSEBSARSQVQVAP
 VAAKRLNPKFPAKGRHEAPLVAWEVTRFMCKKVDLDESLGILVLDYFNPNSTDA
 TDIWKTLRYLDAFGKFFPEKPPNDPVTMTGNTFSGSAVLSDVLDYFNPNSTDA
 PLYLKPLPLMFEGCGLTRFQDGRFPEILIPSTSPSPVPVSKQGAVEEVIQ
 WLTQGHSLVQRQWAPKADGKRPQLRAEPKPIIKERVHFPAETGITFRP
 DVFKTRSVVPAEPEVQRETPEKVSOMLMDLNDNTWOPHLKFSRIQLGLSKTVAI
 MTLPEHQIRHKHTLILSPSGCTGVEGGMESVAKRIDVLGILQSGVQGPGR
 SAKGMVLDVDDTGDEMIETPYQKWECDVDXKHORTLEVRSVASELSASGLNLQL
 LPVLDEDRADKVKRQAIGDLRLINDLRQSEQKHALNPVEFRQWYESSRSLTRV
 SHGRVYPLAGLDPSQEBETFLNMSGDPKKQYQLDIADWLQKRWDTLSKSLNTR
 GRSAVYIMADFWGVLENEVHVGFSKFLDEEBSFTLLSDCVLVARSPAHFPSTIQ
 RVAVFKELLSLDKVIIFSTQGVPLAKSLGGDYDGMNWCWQPEIVDGVNAEM
 PLEPDLRYLAKDKTFLQKAWSHGTSAAKEQITVDMIOKSFHFLQPNFLGMCNTY
 KERLCYINVSNGPETHIDLYKFSIARPAIDKEBAFNMAKKAQDTEGHAFFWDPDLA
 MYKSDSLWGRSGPHIDLYKFSIARPAIDKEBAFNMAKKAQDTEGHAFFWDPDLA
 EYVTFPEISDKSRSSALDTLLKNRIGEVKEYGRVLVQKMERDSDKDPYVVRNVQY
 KWCATITPEAMDKSGANDSVKILLELSFLADREMNTHALLRASTAFKLYYHKSFK
 VQWMAGRQLAVIKAGTMSRPGEGPALMTAFVYAGLMPDKKFTQYVARLEGSGQSEY
 DPVEVYQLGDDDFDGIQIGFTGNTY

ORIGIN

| Query Match | 100.0% | Score | 8045; | DB | 6; | Length | 8045; |
|-----------------------|---------|---|--------|------------|----|--------|-------|
| Best Local Similarity | 100.0%; | Pred. | No. 0; | | | | |
| Matches | 8045; | Conservative | 0; | Mismatches | 0; | Indels | 0; |
| Qy | 1 | GGTACCGGGCCCCCTCGAGGTGCGAGGTATCGATAAGCTTGATCGAATTCCTGAAAA | 60 | | | | |
| Db | 1 | GGTACCGGGCCCCCTCGAGGTGCGAGGTATCGATAAGCTTGATCGAATTCCTGAAAA | 60 | | | | |
| Qy | 61 | GTCTAAAGTCTCCTGTCTCTCATGTAGGTAGCCCTACTGCGTGTGTAGTGTAGTAGGT | 120 | | | | |
| Db | 61 | GTCTAAAGTCTCCTGTCTCTCATGTAGGTAGCCCTACTGCGTGTGTAGTGTAGTAGGT | 120 | | | | |
| Qy | 121 | CCGAATAGGTTGCTACGCTGTCGCCCTTTAGGGGCTGGTTTAAACAGGGGAGACAGGGTACA | 180 | | | | |
| Db | 121 | CCGAATAGGTTGCTACGCTGTCGCCCTTTAGGGGCTGGTTTAAACAGGGGAGACAGGGTACA | 180 | | | | |
| Qy | 181 | GTCTCTCGGGAAACAGAGACGACCAAGCTTTGTGTGTGCATCATCTTTGAGACCATTCGT | 240 | | | | |
| Db | 181 | GTCTCTCGGGAAACAGAGACGACCAAGCTTTGTGTGTGCATCATCTTTGAGACCATTCGT | 240 | | | | |
| Qy | 241 | CTGAGGGTCCATACAAACCATACATGCTGTGTAGAGAGTGGTGGGAATATTAGAGGTTACT | 300 | | | | |
| Db | 241 | CTGAGGGTCCATACAAACCATACATGCTGTGTGTAGAGAGTGGTGGGAATATTAGAGGTTACT | 300 | | | | |
| Qy | 301 | TTGAGAGTTTCTGAAGAGATTGAGACTGTAGTACGTTTCATTTATTCGCTGCGACGAGAA | 360 | | | | |
| Db | 301 | TTGAGAGTTTCTGAAGAGATTGAGACTGTAGTACGTTTCATTTATTCGCTGCGACGAGAA | 360 | | | | |
| Qy | 361 | GGGACGCAATGAGTCGCAATTTGCCAATTTCTGGTTGGCACGCAACATCGATAGTGTACAC | 420 | | | | |
| Db | 361 | GGGACGCAATGAGTCGCAATTTGCCAATTTCTGGTTGGCACGCAACATCGATAGTGTACAC | 420 | | | | |
| Qy | 421 | ATGTGTGCGGCTTGAGACTGTCCAAATGGACCAATGGTAAGCGGTGAGCGCGAGTGGGGTC | 480 | | | | |

Db 1501 GGGTGGGATACAGCCATTTCCAAAATGGTCTCGTGGATCGGAGAAAGAAAATGGA 1560
 Qy 1561 AGGCTACGAGCTGGACACCTGTAGGTCACTGGCTACCTTTCAACACCAACAGTGA 1620
 Db 1561 AGGCTACGAGCTGGACACCTGTAGGTCACTGGCTACCTTTCAACACCAACAGTGA 1620
 Qy 1621 TATCATACCCATACCATCAGGGCCACATACCGCCGCTCTCTCGAAACATCGGGCTGAAT 1680
 Db 1621 TATCATACCCATACCATCAGGGCCACATACCGCCGCTCTCTCGAAACATCGGGCTGAAT 1680
 Qy 1681 GTGACCAATGGCGGACAGCGGGCTTTTCAATCTGTGGGAGCTTTCTTCCACGGGTTT 1740
 Db 1681 GTGACCAATGGCGGACAGCGGGCTTTTCAATCTGTGGGAGCTTTCTTCCACGGGTTT 1740
 Qy 1741 GGGCTGATCGGAGCGTATCGAGCTGGATGGGACTGTTTACCACTCGAATCACTTGAC 1800
 Db 1741 GGGCTGATCGGAGCGTATCGAGCTGGATGGGACTGTTTACCACTCGAATCACTTGAC 1800
 Qy 1801 GGGCTGGGCGGATTTATGGCCATCCACCGGATGATCGGTCCCTTCACTGTTTCCCTGC 1860
 Db 1801 GGGCTGGGCGGATTTATGGCCATCCACCGGATGATCGGTCCCTTCACTGTTTCCCTGC 1860
 Qy 1861 AACCTACTCTCCGACAGATCAGACGAGGCTTTATGCCAGTTTCCGCGGCTTCCCAATT 1920
 Db 1861 AACCTACTCTCCGACAGATCAGACGAGGCTTTATGCCAGTTTCCGCGGCTTCCCAATT 1920
 Qy 1921 ATAGCCACAGACGAAAGCTGCGCAAGTTGTTGCAAAATGCTCAAACTGTTTCAACACTT 1980
 Db 1921 ATAGCCACAGACGAAAGCTGCGCAAGTTGTTGCAAAATGCTCAAACTGTTTCAACACTT 1980
 Qy 1981 TGCAGAAACAGGACCAACACGAAACCACTTGGCCGCTCGCGGAAACCCGCTGCTTT 2040
 Db 1981 TGCAGAAACAGGACCAACACGAAACCACTTGGCCGCTCGCGGAAACCCGCTGCTTT 2040
 Qy 2041 GAACTCGATAGCCCAATTCACCAACAGGTCGCTTTCCCGCTGGCGCTGCTTGTCCC 2100
 Db 2041 GAACTCGATAGCCCAATTCACCAACAGGTCGCTTTCCCGCTGGCGCTGCTTGTCCC 2100
 Qy 2101 GACTTCGTGAAGCAACAGACTGGCGCTATGSCCAAGGCCCATTCATGACAAATGGT 2160
 Db 2101 GACTTCGTGAAGCAACAGACTGGCGCTATGSCCAAGGCCCATTCATGACAAATGGT 2160
 Qy 2161 TAAATCGAGGTCAAAACCTACAAGAGATCAACCGCGCAAGATATACITGTCGTTCAACCC 2220
 Db 2161 TAAATCGAGGTCAAAACCTACAAGAGATCAACCGCGCAAGATATACITGTCGTTCAACCC 2220
 Qy 2221 GACACCCGAAGCTTGAAGCGCAAGCCACAGTGGTGTGTTTATCTCCGACACTTA 2280
 Db 2221 GACACCCGAAGCTTGAAGCGCAAGCCACAGTGGTGTGTTTATCTCCGACACTTA 2280
 Qy 2281 TCGACCTTACGCCCAAGACGCTCCGCTTGGTGAAGCTCGAGTCGGGTAAATCAGGTAT 2340
 Db 2281 TCGACCTTACGCCCAAGACGCTCCGCTTGGTGAAGCTCGAGTCGGGTAAATCAGGTAT 2340
 Qy 2341 ATCTCAGTGGCTTGTCTCTTTTTCGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2400
 Db 2341 ATCTCAGTGGCTTGTCTCTTTTTCGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2400
 Qy 2401 AACCCGTTGTATCTACTCTACAGGCGGATCTCCACGCGCAAAATGAACCTATTAC 2460
 Db 2401 AACCCGTTGTATCTACTCTACAGGCGGATCTCCACGCGCAAAATGAACCTATTAC 2460
 Qy 2461 TCCTAGGAAGAGGAATAGCCCGCTCGAGGAATCATAAACCGGCTCAATTAACGACTACAA 2520
 Db 2461 TCCTAGGAAGAGGAATAGCCCGCTCGAGGAATCATAAACCGGCTCAATTAACGACTACAA 2520
 Qy 2521 CCTGGGCTCAGTGTGTGCGAGACAACTCTCACCCCGACCGCGGAGAGGTGGC 2580
 Db 2521 CCTGGGCTCAGTGTGTGCGAGACAACTCTCACCCCGACCGCGGAGAGGTGGC 2580
 Qy 2581 CGAGGTGACAGGATTTGGTTCGCCATGACAAAGTCTACAGAGCCCTGAACTTTCTTA 2640
 Db 2581 CGAGGTGACAGGATTTGGTTCGCCATGACAAAGTCTACAGAGCCCTGAACTTTCTTA 2640

Qy 2641 CTGGCGAAGATGATCTCCCTGAACAGGCGAGAAAGCAAACTTCTTCAATCGAGGCCAAAGC 2700
 Db 2641 CTGGCGAAGATGATCTCCCTGAACAGGCGAGAAAGCAAACTTCTTCAATCGAGGCCAAAGC 2700
 Qy 2701 TGGAGCTCGAACTGGGTGCCAAAGCCCAAGCCGACCTGTGACACAGCTTCCGTTGTTCCAA 2760
 Db 2701 TGGAGCTCGAACTGGGTGCCAAAGCCCAAGCCGACCTGTGACACAGCTTCCGTTGTTCCAA 2760
 Qy 2761 GGAACCTCCCGCGCGCTTACTGCGCGGCAACAAATGGGCAATTCGACACTGTGTGTCGA 2820
 Db 2761 GGAACCTCCCGCGCGCTTACTGCGCGGCAACAAATGGGCAATTCGACACTGTGTGTCGA 2820
 Qy 2821 GGTGCTTAATAGTGTATGCACTTCCCTCAATACACACAGGTCGAACCTTGGCAGAAC 2880
 Db 2821 GGTGCTTAATAGTGTATGCACTTCCCTCAATACACACAGGTCGAACCTTGGCAGAAC 2880
 Qy 2881 TCTAAGCGGCCAAGTGGCTGAGCGGCCCAACCTTACCAACACCAACGCAAGGATGA 2940
 Db 2881 TCTAAGCGGCCAAGTGGCTGAGCGGCCCAACCTTACCAACACCAACGCAAGGATGA 2940
 Qy 2941 GCCCGCAATGACACTTTCGCTGATCGCGCCCAAGCGCTGCTGCTTCCGCTTCCGACAGG 3000
 Db 2941 GCCCGCAATGACACTTTCGCTGATCGCGCCCAAGCGCTGCTGCTTCCGCTTCCGACAGG 3000
 Qy 3001 TCTCTCTATTACGCGCGCGGATACCCCTTAAAGTTTCCCGGATCCAGTCAATACCGGTTT 3060
 Db 3001 TCTCTCTATTACGCGCGCGGATACCCCTTAAAGTTTCCCGGATCCAGTCAATACCGGTTT 3060
 Qy 3061 CAAAACACCATCTCTGAGAGTGAGAACTCAATCAGTGCACCAAGCGGCGCAAGGCA 3120
 Db 3061 CAAAACACCATCTCTGAGAGTGAGAACTCAATCAGTGCACCAAGCGGCGCAAGGCA 3120
 Qy 3121 GGTGCTGATAATGTTGCGCTGCGCGCCCGCGCTGCTTATTCGAGCGCTTTTGA 3180
 Db 3121 GGTGCTGATAATGTTGCGCTGCGCGCCCGCGCTGCTTATTCGAGCGCTTTTGA 3180
 Qy 3181 CAAAGTACCGACTCGAAGGATGCCAATACGAGAGATCCACGCGGACAGGTCATAGAG 3240
 Db 3181 CAAAGTACCGACTCGAAGGATGCCAATACGAGAGATCCACGCGGACAGGTCATAGAG 3240
 Qy 3241 AGCGGACAGGTGGATTCCTTTGATACATCTCAAGGCACTTCTTATGTTCCAGTGTCTT 3300
 Db 3241 AGCGGACAGGTGGATTCCTTTGATACATCTCAAGGCACTTCTTATGTTCCAGTGTCTT 3300
 Qy 3301 CAGCGCTTCCGCTCAATCAGAGCACTACCCAGAGTGGTGGTGGTGGTGGTGGTGGTGGT 3360
 Db 3301 CAGCGCTTCCGCTCAATCAGAGCACTACCCAGAGTGGTGGTGGTGGTGGTGGTGGTGGT 3360
 Qy 3361 GCCGAGAGAGAGCGGCTGCGATGCCAGCGCTTTGAGGCTGACACTTGTATGAGTC 3420
 Db 3361 GCCGAGAGAGAGCGGCTGCGATGCCAGCGCTTTGAGGCTGACACTTGTATGAGTC 3420
 Qy 3421 TCTTAGCAAGGAAGAAACCAAGTCCCAATAGATAAACCAGCCCTTTCATCGTCTTC 3480
 Db 3421 TCTTAGCAAGGAAGAAACCAAGTCCCAATAGATAAACCAGCCCTTTCATCGTCTTC 3480
 Qy 3481 CCAGGTGAACTTCGTTTCAAGCACTTATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3540
 Db 3481 CCAGGTGAACTTCGTTTCAAGCACTTATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3540
 Qy 3541 CGCAATTTCCCGAGCGAGTCCGCTCAAAATGGACTGCTCGAGCGAAGAAAGCGCTCGATC 3600
 Db 3541 CGCAATTTCCCGAGCGAGTCCGCTCAAAATGGACTGCTCGAGCGAAGAAAGCGCTCGATC 3600
 Qy 3601 TCAGGTCAAGTTCATGCCCGGTTGCTGAGCTGGCTGGCTGAGAAATATTTGGCCGAAAT 3660
 Db 3601 TCAGGTCAAGTTCATGCCCGGTTGCTGAGCTGGCTGGCTGAGAAATATTTGGCCGAAAT 3660
 Qy 3661 TCCCAATGGCTTACACGAGGCTCTCTGCTGTTTGCATGGGAAAGTTTACAGACTCTTTAT 3720
 Db 3661 TCCCAATGGCTTACACGAGGCTCTCTGCTGTTTGCATGGGAAAGTTTACAGACTCTTTAT 3720

QY 3721 GCACGTCAAAAGTAGACTTGGAAAGCAGAGACCTGGGCCCTAAAGTACGACCCCTCTCTGGTC 3780
Db 3721 GCACGTCAAAAGTAGACTTGGAAAGCAGAGACCTGGGCCCTAAAGTACGACCCCTCTCTGGTC 3780
QY 3781 TACCGCGCGAGTGCACAGATATCTGGAAGACTCTCTACCGGCTTGATGCTTTCGGTGG 3840
Db 3781 TACCGCGCGAGTGCACAGATATCTGGAAGACTCTCTACCGGCTTGATGCTTTCGGTGG 3840
QY 3841 TAAACCCCTTCCAGAAAGCCGCCAACGACGCTGTTCTGTAACCGCAATACGCGGCAACTT 3900
Db 3841 TAAACCCCTTCCAGAAAGCCGCCAACGACGCTGTTCTGTAACCGCAATACGCGGCAACTT 3900
QY 3901 TGAGAGCAAGGTAGTGCCTGCTCTCTCTGCTGCTCTAGACTACAATCCGACACATC 3960
Db 3901 TGAGAGCAAGGTAGTGCCTGCTCTCTCTGCTGCTCTAGACTACAATCCGACACATC 3960
QY 3961 GCCTACTGCGCCCTTTACCTTGTGAAGCTGAAAGCCGCTCATGTTTCGAGCAGGCTGTGC 4020
Db 3961 GCCTACTGCGCCCTTTACCTTGTGAAGCTGAAAGCCGCTCATGTTTCGAGCAGGCTGTGC 4020
QY 4021 ACTCACCCGTCGGTTCGGTCTCTGATAGTCTTTCGAGATCCCTATACCGTTCGCTACGAG 4080
Db 4021 ACTCACCCGTCGGTTCGGTCTCTGATAGTCTTTCGAGATCCCTATACCGTTCGCTACGAG 4080
QY 4081 CACAGCCCCAAGTGTACCGCGGTGTCAGCAAAACCAACAGAGTCGGTTCGAAAGATCAT 4140
Db 4081 CACAGCCCCAAGTGTACCGCGGTGTCAGCAAAACCAACAGAGTCGGTTCGAAAGATCAT 4140
QY 4141 CCAGTGCCTCAGATGGGGCAATCTCTCTGTTAGCGCGCAATGGCGGCTTCTTCGC 4200
Db 4141 CCAGTGCCTCAGATGGGGCAATCTCTCTGTTAGCGCGCAATGGCGGCTTCTTCGC 4200
QY 4201 CAAAGATGCGCGGATACAGGAAACCTCTCAGGAGTTCAGAGTCCGCGCGGAGGACCGGAA 4260
Db 4201 CAAAGATGCGCGGATACAGGAAACCTCTCAGGAGTTCAGAGTCCGCGCGGAGGACCGGAA 4260
QY 4261 ACCATCATCAAGAGAGAGTCCACTCTTTCGCGAGACCGGATTAATGTTCCGACCTGA 4320
Db 4261 ACCATCATCAAGAGAGAGTCCACTCTTTCGCGAGACCGGATTAATGTTCCGACCTGA 4320
QY 4321 TGTGTTCAAGCAGAGATCTGCTCTCGGCGAGAGAACTCTGAGAGCAACGAGACCGAGTT 4380
Db 4321 TGTGTTCAAGCAGAGATCTGCTCTCGGCGAGAGAACTCTGAGAGCAACGAGACCGAGTT 4380
QY 4381 CAAAGTTAGTCAAAATGCTGACTGCTCTGCAACTCGACAAACACTTGGCAGCCGCA 4440
Db 4381 CAAAGTTAGTCAAAATGCTGACTGCTCTGCAACTCGACAAACACTTGGCAGCCGCA 4440
QY 4441 CCTCAAGTTGTTCTCCGCTATCCAGCTCGGTCTGAGTAGACATATGCCATTAACAT 4500
Db 4441 CCTCAAGTTGTTCTCCGCTATCCAGCTCGGTCTGAGTAGACATATGCCATTAACAT 4500
QY 4501 GGAGCCTCACAGATCAGACACCAAGACCGATCTTCTTTCACCTTCAGGCACTGGGGA 4560
Db 4501 GGAGCCTCACAGATCAGACACCAAGACCGATCTTCTTTCACCTTCAGGCACTGGGGA 4560
QY 4561 AGTGATGATAGCGGTGAGGCGGATGTCGGAGCGTGGCCAGCGGTTGGTTTCGGCCAAAGGAAT 4620
Db 4561 AGTGATGATAGCGGTGAGGCGGATGTCGGAGCGTGGCCAGCGGTTGGTTTCGGCCAAAGGAAT 4620
QY 4621 TCTCGGTTTGGGTGATGTCCTCTGCTGTCGCAAGGCGGTTGGTTTCGGCCAAAGGAAT 4680
Db 4621 TCTCGGTTTGGGTGATGTCCTCTGCTGTCGCAAGGCGGTTGGTTTCGGCCAAAGGAAT 4680
QY 4681 GTGCGTTATCGAGTTGACGACACAGCGGATGAGGATTCGAGACATACCCCTCCCA 4740
Db 4681 GTGCGTTATCGAGTTGACGACACAGCGGATGAGGATTCGAGACATACCCCTCCCA 4740
QY 4741 GCSCAAGTGGGAATGCGACTTCTGTTGATAAACATCAACGTACCTTCGAAGTCCGAGCGT 4800
Db 4741 GCSCAAGTGGGAATGCGACTTCTGTTGATAAACATCAACGTACCTTCGAAGTCCGAGCGT 4800
QY 4801 GGCTTCTGAAGTACGAGTGGTCTCAACCTACAGCTGTTACTGCTGGAAGATAG 4860

Db 4801 GGCTTCTGAAGTACGAGTGGTCTCAACCTACAGCTGTACCTCTCTGGAAGATAG 4860
QY 4861 AGCCAGGGAACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCA 4920
Db 4861 AGCCAGGGAACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCA 4920
QY 4921 ACAGCAGTTACGAGGACAAAGCAATGCTTTGAATGCGCCAGTGGAAATTCGCCAATGGGT 4980
Db 4921 ACAGCAGTTACGAGGACAAAGCAATGCTTTGAATGCGCCAGTGGAAATTCGCCAATGGGT 4980
QY 4981 TTACGAGAGTTATTCGAGTCCGCAACTCGAGTCAGCCAGCGCGGTGCTTTCCTTTC 5040
Db 4981 TTACGAGAGTTATTCGAGTCCGCAACTCGAGTCAGCCAGCGCGGTGCTTTCCTTTC 5040
QY 5041 TGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCC 5100
Db 5041 TGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCC 5100
QY 5101 CAAGAAGCAAAAGTACTTTCAGAGACATCGCTCGGATCTTCAAAAGCGGAATGTGACAC 5160
Db 5101 CAAGAAGCAAAAGTACTTTCAGAGACATCGCTCGGATCTTCAAAAGCGGAATGTGACAC 5160
QY 5161 GTTGAAGTCCAAAGCTGAACATCCGTCGCTCGATCAGCATACATTTACATGATTCGCGA 5220
Db 5161 GTTGAAGTCCAAAGCTGAACATCCGTCGCTCGATCAGCATACATTTACATGATTCGCGA 5220
QY 5221 TTTCTGCGGTGCTTCAGGAAATGAGTTTCATGTCGGATCTTCTCAAAAGTTCAGAGA 5280
Db 5221 TTTCTGCGGTGCTTCAGGAAATGAGTTTCATGTCGGATCTTCTCAAAAGTTCAGAGA 5280
QY 5281 CGAGGAGAGTCTTTTACACTCTTATCGGACTGTGATGTCCTGTCGCGGATCCCCAGC 5340
Db 5281 CGAGGAGAGTCTTTTACACTCTTATCGGACTGTGATGTCCTGTCGCGGATCCCCAGC 5340
QY 5341 CCATTTCCCTAGTGAATCCAAAGGAGATGTACCGCTTGTCTAAGAGATATCTGG 5400
Db 5341 CCATTTCCCTAGTGAATCCAAAGGAGATGTACCGCTTGTCTAAGAGATATCTGG 5400
QY 5401 CAAGGATGTAATCATCTTCTACTAAGAGAGATGTACCGCTTGTCTAAGAGATATCTGG 5460
Db 5401 CAAGGATGTAATCATCTTCTACTAAGAGAGATGTACCGCTTGTCTAAGAGATATCTGG 5460
QY 5461 TGGAGACTA CGAGCGGATATGCGCTGGGTCTCTGGGATCCGAGATTCGATGGTTT 5520
Db 5461 TGGAGACTA CGAGCGGATATGCGCTGGGTCTCTGGGATCCGAGATTCGATGGTTT 5520
QY 5521 CGTCAATCGGAAATGCTCTGAGCGCGGATCTAGTACTTAAGAGAGCAAAAC 5580
Db 5521 CGTCAATCGGAAATGCTCTGAGCGCGGATCTAGTACTTAAGAGAGCAAAAC 5580
QY 5581 GACTTTCAAAACAACTTATGGCTCACAAGGACCGGCTCAGCGGCCAAAGAGCAGACTAC 5640
Db 5581 GACTTTCAAAACAACTTATGGCTCACAAGGACCGGCTCAGCGGCCAAAGAGCAGACTAC 5640
QY 5641 ATAGCATATGATCCAGAGAGCTTCCATTTGCGCTTCGAGCCCAACTTCTTGGGCAATG 5700
Db 5641 ATAGCATATGATCCAGAGAGCTTCCATTTGCGCTTCGAGCCCAACTTCTTGGGCAATG 5700
QY 5701 CACTAACTCAAAAGAAAGGCTCTGTTTACATCAACATAGTGTCTTAAACAAGCGGCAAT 5760
Db 5701 CACTAACTCAAAAGAAAGGCTCTGTTTACATCAACATAGTGTCTTAAACAAGCGGCAAT 5760
QY 5761 CATTTCTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGATTTCTTTTAA 5820
Db 5761 CATTTCTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGATTTCTTTTAA 5820
QY 5821 CGAAGCAAGTGGGCTCAATTCGCTAGGGAACCTGCTTGGCGGTGCAATTTCTTCTCTGA 5880
Db 5821 CGAAGCAAGTGGGCTCAATTCGCTAGGGAACCTGCTTGGCGGTGCAATTTCTTCTCTGA 5880
QY 5881 CCATATGTACAAAGAGCAGTGGCTCGGCGCGGAGAGCTTACCACATTTATTCACTA 5940

Db 5881 CCCAATGTACAAGAGCGACAGTTGGCTCGGGCGGAGAGCGCTACCCACATTAATTGACTA 5940
Qy 5941 CCTGAAATCTCCATCGCGAGCGCTGGATTTGACAAAGAACTGGAAGCCTTCCCAATGC 6000
Db 5941 CCTGAAATCTCCATCGCGAGCGCTGGATTTGACAAAGAACTGGAAGCCTTCCCAATGC 6000
Qy 6001 CATGAAAGCGCCCAAGAGTACAGAAAGACGGGGCTCACTTTTGGGATCCGGATCTCGCTTC 6060
Db 6001 CATGAAAGCGCCCAAGAGTACAGAAAGACGGGGCTCACTTTTGGGATCCGGATCTCGCTTC 6060
Qy 6061 CTACTACAGCTTCTTCAAGGAGATTAGCGACAAGTCGGGATCGTCCGACTGCTATTAC 6120
Db 6061 CTACTACAGCTTCTTCAAGGAGATTAGCGACAAGTCGGGATCGTCCGACTGCTATTAC 6120
Qy 6121 GACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAAATATGGCAGGTTGGTCAAAAACAA 6180
Db 6121 GACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAAATATGGCAGGTTGGTCAAAAACAA 6180
Qy 6181 GGAGATGAGAGACAGCAAGAACCCCTACCTCTGTCGGGTCAACAGAGTTTATGAAAATG 6240
Db 6181 GGAGATGAGAGACAGCAAGAACCCCTACCTCTGTCGGGTCAACAGAGTTTATGAAAATG 6240
Qy 6241 GTGGCCATCAGCCTGAGGCGATGGACAAATCCGGAGCAAAATATGATCTAAGGTGAT 6300
Db 6241 GTGGCCATCAGCCTGAGGCGATGGACAAATCCGGAGCAAAATATGATCTAAGGTGAT 6300
Qy 6301 CAGGTTGCTGAGGTGCTCTTCTCGGGAACCGTGAGATGAATACATGGSCATTTGCTGAG 6360
Db 6301 CAGGTTGCTGAGGTGCTCTTCTCGGGAACCGTGAGATGAATACATGGSCATTTGCTGAG 6360
Qy 6361 GGCTAGCAGCGCTTTCAGCTGTACTACCAAGAGCCCCAAGTTCTGTGTGGCAGATGGC 6420
Db 6361 GGCTAGCAGCGCTTTCAGCTGTACTACCAAGAGCCCCAAGTTCTGTGTGGCAGATGGC 6420
Qy 6421 GGGCAGACAGCTCGCGTACATTAAGGGCGAGATGACAGAGCAGACCCGGTGAAGCGCCCC 6480
Db 6421 GGGCAGACAGCTCGCGTACATTAAGGGCGAGATGACAGAGCAGACCCGGTGAAGCGCCCC 6480
Qy 6481 GGGCTTGATGACCGGCTCATGTATGGGGCTTGATCGCGGATGAAGTTTACGAGCA 6540
Db 6481 GGGCTTGATGACCGGCTCATGTATGGGGCTTGATCGCGGATGAAGTTTACGAGCA 6540
Qy 6541 GTATGTGCCAGGCTGGAGGCGGATGGATCGGAGTACCTGATCCGGAGGTCTATGAAT 6600
Db 6541 GTATGTGCCAGGCTGGAGGCGGATGGATCGGAGTACCTGATCCGGAGGTCTATGAAT 6600
Qy 6601 GCTGGCGATGATGATTTTGAATGGATTTGTTTTCAGAGGAATGGCGATTTAGTCGGG 6660
Db 6601 GCTGGCGATGATGATTTTGAATGGATTTGTTTTCAGAGGAATGGCGATTTAGTCGGG 6660
Qy 6661 TTGCTTGAGGGGCGCTGCATCTTATTTTCTGATTTTCCGAGTCTTTGGTGTGGG 6720
Db 6661 TTGCTTGAGGGGCGCTGCATCTTATTTTCTGATTTTCCGAGTCTTTGGTGTGGG 6720
Qy 6721 AATGGGAATCTTGAACATGTTCCCTATGGGTGAGTGTCTCATTTATGGACTTCAGTTT 6780
Db 6721 AATGGGAATCTTGAACATGTTCCCTATGGGTGAGTGTCTCATTTATGGACTTCAGTTT 6780
Qy 6781 TGGAGTTTCTTTCGGCTTTCAGATCTGCTGACAGAGCGAGCGCTGGACATTCGGTTG 6840
Db 6781 TGGAGTTTCTTTCGGCTTTCAGATCTGCTGACAGAGCGAGCGCTGGACATTCGGTTG 6840
Qy 6841 GTCCGTAGAACATAAAACGAAATGGGAGATGATGATTTCTTCTGGCGCTGATAATTT 6900
Db 6841 GTCCGTAGAACATAAAACGAAATGGGAGATGATGATTTCTTCTGGCGCTGATAATTT 6900
Qy 6901 GAGATATTTGGTGTGCTGCTTGTATCCATGTCGAGATATTCGTTCTTACTTGTGCC 6960
Db 6901 GAGATATTTGGTGTGCTGCTTGTATCCATGTCGAGATATTCGTTCTTACTTGTGCC 6960
Qy 6961 CGCGAAGCAGATTTGGTCCAACTTTCAAACAATAGTCAGCAACCTTCAGGAAACTATCGA 7020
Db 6961 CGCGAAGCAGATTTGGTCCAACTTTCAAACAATAGTCAGCAACCTTCAGGAAACTATCGA 7020

Qy 7021 TCTCGATCAGAGCGCTTAGAATAAATAGGTCCTCGCTCCCTTAGCTCCGAGCTCCATC 7080
Db 7021 TCTCGATCAGAGCGCTTAGAATAAATAGGTCCTCGCTCCCTTAGCTCCGAGCTCCATC 7080
Qy 7081 CTAGGTGGAGCACC CGGGATCTCTAGACTCTCTTTTTCAGACCTGTCTAAGTGAGATATA 7140
Db 7081 CTAGGTGGAGCACC CGGGATCTCTCTAGACTCTCTTTTTCAGACCTGTCTAAGTGAGATATA 7140
Qy 7141 GTGCAAGCAGTGGTGTGTACACTTTGGTATTACATCAACAGACAAATTTGAGATAGCT 7200
Db 7141 GTGCAAGCAGTGGTGTGTGTACACTTTGGTATTACATCAACAGACAAATTTGAGATAGCT 7200
Qy 7201 CATTCCTCCGGTTTCTTCGATTGCGACTCTATTGTATGATTTTTCATCTACCC 7260
Db 7201 CATTCCTCCGGTTTCTTCGATTGCGACTCTATTGTATGATTTTTCATCTACCC 7260
Qy 7261 CTGCAAGCATAACTCTAGCAAAATCTTTTGTGTAATGAATAGGCTCTTTTGGGATTC 7320
Db 7261 CTGCAAGCATAACTCTAGCAAAATCTTTTGTGTAATGAATAGGCTCTTTTGGGATTC 7320
Qy 7321 GAGGTTTTCGACCTTGAAACCTTTTGATCTTTTCTTTTTCGTTTGGTTGGAAA 7380
Db 7321 GAGGTTTTCGACCTTGAAACCTTTTGATCTTTTCTTTTTCGTTTGGTTGGAAA 7380
Qy 7381 GTATATCAACTCTA OCTGATCACTCGGAAGCAACAAACAAACAAACAAACAAACA 7440
Db 7381 GTATATCAACTCTA OCTGATCACTCGGAAGCAACAAACAAACAAACAAACAAACA 7440
Qy 7441 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7500
Db 7441 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7500
Qy 7501 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7560
Db 7501 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7560
Qy 7561 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7620
Db 7561 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7620
Qy 7621 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7680
Db 7621 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7680
Qy 7681 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7740
Db 7681 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 7740
Qy 7741 ACGGAGCGACCTCATAGCTTTATGGGAAATTAATACAGTCCAGAGATACATAACAAATAC 7800
Db 7741 ACGGAGCGACCTCATAGCTTTATGGGAAATTAATACAGTCCAGAGATACATAACAAATAC 7800
Qy 7801 AGCAACAAATGAAGTTCAAGAGTTTCTCGCCGCGGTTCCACCAACAGCAGCCAC 7860
Db 7801 AGCAACAAATGAAGTTCAAGAGTTTCTCGCCGCGGTTCCACCAACAGCAGCCAC 7860
Qy 7861 CACCCACACATGAAGTTCAAGAGTTTCTCGCCGCGGTTCCACCAACAGCAGCCAC 7920
Db 7861 CACCCACACATGAAGTTCAAGAGTTTCTCGCCGCGGTTCCACCAACAGCAGCCAC 7920
Qy 7921 ATCCCTATAGTCGATGATTTATGGCGCGGAAATTCGATATCAAGCTTATCGATACCGT 7980
Db 7921 ATCCCTATAGTCGATGATTTATGGCGCGGAAATTCGATATCAAGCTTATCGATACCGT 7980
Qy 7981 CGACCTCGAGGGGGGGCGGTTACCAATTCGCCCTATAGTGAGTCGTTTACAATCCCA 8040
Db 7981 CGACCTCGAGGGGGGGCGGTTACCAATTCGCCCTATAGTGAGTCGTTTACAATCCCA 8040
Qy 8041 GTCAC 8045
Db 8041 GTCAC 8045

[illegible]

ORIGIN

Query Match 52.3%; Score 4206; DB 8; Length 4206;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2447 ATGAACCCCTATTACTCTAGGAAGAGGAATAGCCCGCTCGAGGAAATCATTAACCCGGCTC 2506


```
/product="similar to Arabidopsis thaliana (Mouse-ear
cress). Putative UDP-glucose:glycoprotein
glucosyltransferase, 101200- 91134"
/protein_id="AA08766.2"
/db_xref="GI:28829249"
/translation="MARIFEFVLLIIVFISNVLVESGNDNSPSKSIQLSVLSN
WGPSTVLEAEFLHNDKSLFKFIEFNKIDFTSNYSOKIYVESITSLMKVLSLN
TQFSEFISDLNRTYSPVETRYQLAISNMKLNIEHSITTAADNTIILPNSGGWY
KIKIILIDVNEISLPSKVAVDDENEFIRLYDFDHIPTLIANTVSSSSPSPI
PILVYDIKSEFFKLHPKLQFSQMKIKICLYRVVQSSKNLQOGYSELISKN
LEYKMDASAKLKKIIDIGVSKTIININEDVQGFNFKLQKRPKLTSLKSTFRSY
LMARQSEAKIKELQKIQSAKIIQSDGPLRSEIYSQKFTLSNLSKITLNE
SLKSVIEGNOKIIPSTTQDTLLNGLRIDNDELPSIRIELRIEVEHSTIIOOQGPL
SSKTQDIISQALPRTOLLPTKEELANGNEPFSVLNLELDIYVROBEPKQSSV
LDKPVTSQDPIFKNLLTIVILDMNNINTEIIPRIQEWQNSILPRIOLIENF
KSNNNNNNNNNQNSSTNFIQKOLKAKVFLIKNSLNGNCGAFFIATLYFFKM
YINELGIRSVLSVPAVLQOMGVSRSIQHALTNTDFNLLESNNQILIERLELD
TTTQSTYTTITILKLPFVNGVQVYQVSNIDQSLFLLVSLNDEFNKLPLFKESIL
STTQATYTTITILKLPFVNGVQVYQVSNIDQSLFLLVSLNDEFNKLPLFKESIL
NLLVRNNENDEONLNLIVIGPDHYNTRDIISLELLQEKELCKNCKLTFTSNPI
DINSVNTAGNEOILGLIITLKHGKILTPQIVIGLFEKQSDPTIIDSFTMKOI
IELSGFDIAANDIWAQSVNLPKSSVKCYKQYLGIOSTNKSPIANGLIIPPLSY
DEASFIQDFKLLLEIMIRAKKIFELNSDPILKOKSNLKSIDLLNKVQSVLVGYI
NNGNGLDNIKKIRIPNSISFSGKPTLSSSSSSSSSSSNSNDVPLFLMIPNFVK
SOKLVPMVREFTSKNLINPDVILNPPVLSLPLKTYTYVILKSSEFNENLVNQF
LEGANDMTNAPAGLELLNPISTQNTKQDTIIVMNFYQYOLKNGPQWIKLTIAPI
GRSSDIMDVHPNOKETEVIVPHRLAVIDSIVQSLSVSVYRKAGQELRPLQPI
DEYKQEQEOKLQKNSGFFSNLPSSKNDATDSVATQKSNLDTIHFVSAGSH
LYERFKLMLSVKNTPESIPKFWFLKNLYSPAKKEPIPEWAKYGFQYELVYKWPW
MLRQTEIKYKILFLDVLDFDVPKLIIFVDADQVVRDIDJELWMDLHGLASLG
YTPFCDNSKDEGFRWFGYRQHLRAGRSYHISALYVVDLFRFRRLAAGDQURATYD
QLSRDNLKLDLDDLPYLYVRIHSLPQEWLWCTWCQDESKSAKTIIDCNPLD
TKTKLENAVRIIDETWTDNEAKEFELKIDQSKHHRQIELDHQNLQPNKPIENIDD
ILLNLEAQKOLF"
complement(join(10993..12657,12738..12853,13014..13262,
13781..13814,14302..14529))
/notes="Geneid exon scores (in order of location ranges):
175.98, 4.12, 21.10, -3.61, 20.21 - GSCJ_ID dd_01880"
/codon_start=1
/product="similar to Neurospora crassa. Probable branching
enzyme (bel)"
/protein_id="AA08768.2"
/db_xref="GI:28823250"
/translation="MNVDIADIIESENDOENMQSDNSOSPILKEDHNKIKE
FEIIFENKISNLKIEISKNEIETLSQIFSIIVKLIRDYSKATDGTQVIHDD
PWEIPEYKIVIKRHNQVNTIQKLEESRSGLLKFSQGYEFGVTKDGYNIREWLP
AHEVYLVGDFPNWNTSPLEBDNYGRMSIPIPNNSGECALPHGSKIYLVKLANG
FDYRTPANKVEQTKENPVDGVFNPSKQYVFNKSPMKPTELRIYEAHVGMSEL
PEIYSYKFKETVLPMKELCYNCIQLMAVMEHAYASFGYQVTFPAISRRGTPEE
LKEMIDRAHEGGLLVFLDVVSHASKNVLDGLNQLDGDTHHIFHSGGRGHELDWDR
FNYGNWVVRFLLSNRYVDEYHDFGFRFDGVTSMIYTHHGLSPACSYDDYFGGVDI
EDALNYLTANVMYTLNRPSTVIAEVTGLATCRPFEGGDFDYRLAMGIPDRVI
ELVKEKEDNMNGTIAHMLSNRRYKEKNIYAESHDQSLVGDITFLWLMKEMYN
NSVTEETPIIDRGSLHMKRLITSLIGDGYLNFMGNEFHPWEDVFPREGNNSL
HHAEREDLYNRLRYQLRDFDIANNKAEQERFLWSSDPAYISLHEDDKIIVPER
ASLIFENFHPKSPSDYRIGSGVPGKEFNVLDSDRKEFGGHVIGKDNHYIEDXEW
HDKYSLIILYIPSRCLVLKKVD"
join(15053..15120,15221..15743)
/notes="Geneid exon scores (in order of location ranges):
-0.56, 37.71 - GSCJ_ID dd_01882"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA08769.2"
/db_xref="GI:28823251"
/translation="MGQSKGELQYPENKPLFNVKIEDCKIYDTKDDTFLSLQYND
NNKQAKNSDSDDDDDSDSSSTTTTTRRNVNNSKEPIKTSSTNGVQSGLESL
LNNCTNDSDDDDDDSDDEEDNDNSFENNPNLTIEEEKIKIFQEKYKQOQK
KLKANKKIRKQPIYWFQISPTSLKQSQSYFKQK"
complement(join(16609..16729,16860..16934,17034..17080,
17169..17177))
/notes="Geneid exon scores (in order of location ranges):
6.96, 13.18, -2.05, 0.27 - GSCJ_ID dd_01883"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA051794.1"
/db_xref="GI:28829252"
/translation="MAGPWTIEPKNGHGIPTWAPRSTTVHVQVVGGRVGRVMOAG
AKSPDENVGETSPSRGSPGSRNLNIGSRTLKVMTA"
complement(19502..21007)
/notes="Geneid exon scores (in order of location ranges):
108.08 - GSCJ_ID dd_01875"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Futa (Fragment)"
/protein_id="AA051795.1"
/db_xref="GI:28829253"
/translation="MKSLRNSVSIILKIVLITSTCLLLINIALIYEDPNLIDSNCKTT
TITITQNNKNGKLENLNSKKKLLILDVLPNEDTVMSSECTSTTNOQCGFT
TDSILANESGLIYIPRTTKYNNRSTSVHRTSEKQLFFQWMSHESMYPLENDTI
HUSLNYTIPGLNDYQKAGVSDDEEYKDIQSHIYTPVGVFPFGSGHSYSLF
KEGLKVEIPKGRDSIMMAGKRNQKLGRTLVKQMMRFTTIDYSGKLNKNDLKVN
YGMVLLRSLQKILRKDYFVAIENSCEDYVTEKLWEPLSVETIPLYLATIDE
FLPHDAIINIANFNSINHLMKYVDVGLNETLREKLAWIKKPLPKPKFNLLNOSLN
NONLPCSIKLVNLNDNNINVTPLSKPKOCIEGKFTLPITYTEEKERGLLRLOK
QQEQKLELQNNNDYTNDEEESNDNSGNSGNSGNSGNSNSI"
complement(join(21960..26919,26946..27251))
/notes="Geneid exon scores (in order of location ranges):
530.44, 21.78 - GSCJ_ID dd_03196"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Nucleotide exchange factor RasGEF R"
/protein_id="AA051796.1"
/db_xref="GI:28829254"
/translation="MKBEIANGPQEVNPLYLEGINKKIIIEKTVANIDTADKTDNL
QNLIESLALHNINLLYFVDFKPIISYDQIVHYSHFKEVVKIIFQVCQDLIKAAQ
ALINSCDYLTTSNFKAGREAIKTKTMTATCSNFELKYNPDLEVDLHQHQ
AQEOEQOKEQOYQEOBOLIKOLKEQLKQEOOQOQOQOQOQOQOQOQOQOQOQ
EKEKNDKEKEKLLKOLLAEQKQKELEKQOQOQOQOQOQOQOQOQOQOQOQOQ
QASDPILIPYILKTSQSDLLSGLSPKVNSPGSNLCKFTESAERVAERDGTANI
LNNKIKELSLGSDSEAPETVAARIISNAIISIKELRYKITGLTNLSEHVSIVQV
ARLAIKNKSDPYQOLELAIEKFNDMTKRIIYVKSISKSNMNRKQFVDSLSGRS
PVVPEKSIPTSTSERILSHIKNNNNNNNNNNSTNNLQFETPSLSNHS
QTNQOPLQSPILLINQLQSTSSSSSSSSSSSLNSIQLPOATTTTTTATPSTPS
TSTSTSPNSLSISDQQLRKQKHQIVIPKVDKSVANQOQOQOQOQOQOQOQOQOQ
NNESSVTAATTTTTTSTTASITNNVTINPNKOLPATPTTGTSTSTPTPTPTST
SHSTKNNNSPETSFGVESPKITGKLSFKFVKETPLGFDSSNSSSPSNNSTNSS
SONDKONNNKENFVDKOKTIGKLSFKFVKETPLGFDSSNSSSPSNNSTNSS
SSSTKNNNSPETSFGVESPKITGKLSFKFVKETPLGFDSSNSSSPSNNSTNSS
TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
RSFTTIGLVSGKHSVVVETPKSKLQFHESAKQILSIISMNPFSEKQFQMNSEVIS
```


[illegible]

[illegible]

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, C., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Digar, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregor, S., Guadalupe, R., Guevara, W., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamal, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.B., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hui, Y., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwar, L., Louise, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Margum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Marwinsky, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwono, G., Olarunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pflanz, K., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soja, J., Steimle, M., Strong, R., Sutton, A., Tatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, U., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 118689)
Worley, K.C.

Direct Submission
Submitted (09-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 118689)
Worley, K.C.

Direct Submission
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KDZH
Center clone name: CH230-457B19
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 101909 bases at least Q40
Consensus quality: 106962 bases at least Q30

Consensus quality: 114915 bases at least Q20
Estimated insert size: 98018; sum-of-contigs estimation
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1487: contig of 1487 bp in length
* 1488: gap of unknown length
* 1489: contig of 1320 bp in length
* 1490: gap of unknown length
* 1491: contig of 1280 bp in length
* 1492: gap of unknown length
* 1493: contig of 1003 bp in length
* 1494: gap of unknown length
* 1495: contig of 1219 bp in length
* 1496: gap of unknown length
* 1497: contig of 1176 bp in length
* 1498: gap of unknown length
* 1499: contig of 1587 bp in length
* 1500: gap of unknown length
* 1501: contig of 1238 bp in length
* 1502: gap of unknown length
* 1503: contig of 1057 bp in length
* 1504: gap of unknown length
* 1505: contig of 1593 bp in length
* 1506: gap of unknown length
* 1507: contig of 1122 bp in length
* 1508: gap of unknown length
* 1509: contig of 1072 bp in length
* 1510: gap of unknown length
* 1511: contig of 1054 bp in length
* 1512: gap of unknown length
* 1513: contig of 1397 bp in length
* 1514: gap of unknown length
* 1515: contig of 1472 bp in length
* 1516: gap of unknown length
* 1517: contig of 1176 bp in length
* 1518: gap of unknown length
* 1519: contig of 1498 bp in length
* 1520: gap of unknown length
* 1521: contig of 1287 bp in length
* 1522: gap of unknown length
* 1523: contig of 1074 bp in length
* 1524: gap of unknown length
* 1525: contig of 1117 bp in length
* 1526: gap of unknown length
* 1527: contig of 1529 bp in length
* 1528: gap of unknown length
* 1529: contig of 1718 bp in length
* 1530: gap of unknown length
* 1531: contig of 1127 bp in length
* 1532: gap of unknown length
* 1533: contig of 1000 bp in length
* 1534: gap of unknown length
* 1535: contig of 1215 bp in length
* 1536: gap of unknown length
* 1537: contig of 1304 bp in length
* 1538: gap of unknown length
* 1539: contig of 1431 bp in length
* 1540: gap of unknown length
* 1541: contig of 1739 bp in length
* 1542: gap of unknown length
* 1543: contig of 1640 bp in length
* 1544: gap of unknown length
* 1545: contig of 1294 bp in length
* 1546: gap of unknown length
* 1547: contig of 1294 bp in length
* 1548: gap of unknown length
* 1549: contig of 1294 bp in length
* 1550: gap of unknown length
* 1551: contig of 1294 bp in length
* 1552: gap of unknown length
* 1553: contig of 1294 bp in length
* 1554: gap of unknown length
* 1555: contig of 1294 bp in length
* 1556: gap of unknown length
* 1557: contig of 1294 bp in length
* 1558: gap of unknown length
* 1559: contig of 1294 bp in length
* 1560: gap of unknown length
* 1561: contig of 1294 bp in length
* 1562: gap of unknown length
* 1563: contig of 1294 bp in length
* 1564: gap of unknown length
* 1565: contig of 1294 bp in length
* 1566: gap of unknown length
* 1567: contig of 1294 bp in length
* 1568: gap of unknown length
* 1569: contig of 1294 bp in length
* 1570: gap of unknown length
* 1571: contig of 1294 bp in length
* 1572: gap of unknown length
* 1573: contig of 1294 bp in length
* 1574: gap of unknown length
* 1575: contig of 1294 bp in length
* 1576: gap of unknown length
* 1577: contig of 1294 bp in length
* 1578: gap of unknown length
* 1579: contig of 1294 bp in length
* 1580: gap of unknown length
* 1581: contig of 1294 bp in length
* 1582: gap of unknown length
* 1583: contig of 1294 bp in length
* 1584: gap of unknown length
* 1585: contig of 1294 bp in length
* 1586: gap of unknown length
* 1587: contig of 1294 bp in length
* 1588: gap of unknown length
* 1589: contig of 1294 bp in length
* 1590: gap of unknown length
* 1591: contig of 1294 bp in length
* 1592: gap of unknown length
* 1593: contig of 1294 bp in length
* 1594: gap of unknown length
* 1595: contig of 1294 bp in length
* 1596: gap of unknown length
* 1597: contig of 1294 bp in length
* 1598: gap of unknown length
* 1599: contig of 1294 bp in length
* 1600: gap of unknown length
* 1601: contig of 1294 bp in length
* 1602: gap of unknown length
* 1603: contig of 1294 bp in length
* 1604: gap of unknown length
* 1605: contig of 1294 bp in length
* 1606: gap of unknown length
* 1607: contig of 1294 bp in length
* 1608: gap of unknown length
* 1609: contig of 1294 bp in length
* 1610: gap of unknown length
* 1611: contig of 1294 bp in length
* 1612: gap of unknown length
* 1613: contig of 1294 bp in length
* 1614: gap of unknown length
* 1615: contig of 1294 bp in length
* 1616: gap of unknown length
* 1617: contig of 1294 bp in length
* 1618: gap of unknown length
* 1619: contig of 1294 bp in length
* 1620: gap of unknown length
* 1621: contig of 1294 bp in length
* 1622: gap of unknown length
* 1623: contig of 1294 bp in length
* 1624: gap of unknown length
* 1625: contig of 1294 bp in length
* 1626: gap of unknown length
* 1627: contig of 1294 bp in length
* 1628: gap of unknown length
* 1629: contig of 1294 bp in length
* 1630: gap of unknown length
* 1631: contig of 1294 bp in length
* 1632: gap of unknown length
* 1633: contig of 1294 bp in length
* 1634: gap of unknown length
* 1635: contig of 1294 bp in length
* 1636: gap of unknown length
* 1637: contig of 1294 bp in length
* 1638: gap of unknown length
* 1639: contig of 1294 bp in length
* 1640: gap of unknown length
* 1641: contig of 1294 bp in length
* 1642: gap of unknown length
* 1643: contig of 1294 bp in length
* 1644: gap of unknown length
* 1645: contig of 1294 bp in length
* 1646: gap of unknown length
* 1647: contig of 1294 bp in length
* 1648: gap of unknown length
* 1649: contig of 1294 bp in length
* 1650: gap of unknown length
* 1651: contig of 1294 bp in length
* 1652: gap of unknown length
* 1653: contig of 1294 bp in length
* 1654: gap of unknown length
* 1655: contig of 1294 bp in length
* 1656: gap of unknown length
* 1657: contig of 1294 bp in length
* 1658: gap of unknown length
* 1659: contig of 1294 bp in length
* 1660: gap of unknown length
* 1661: contig of 1294 bp in length
* 1662: gap of unknown length
* 1663: contig of 1294 bp in length
* 1664: gap of unknown length
* 1665: contig of 1294 bp in length
* 1666: gap of unknown length
* 1667: contig of 1294 bp in length
* 1668: gap of unknown length
* 1669: contig of 1294 bp in length
* 1670: gap of unknown length
* 1671: contig of 1294 bp in length
* 1672: gap of unknown length
* 1673: contig of 1294 bp in length
* 1674: gap of unknown length
* 1675: contig of 1294 bp in length
* 1676: gap of unknown length
* 1677: contig of 1294 bp in length
* 1678: gap of unknown length
* 1679: contig of 1294 bp in length
* 1680: gap of unknown length
* 1681: contig of 1294 bp in length
* 1682: gap of unknown length
* 1683: contig of 1294 bp in length
* 1684: gap of unknown length
* 1685: contig of 1294 bp in length
* 1686: gap of unknown length
* 1687: contig of 1294 bp in length
* 1688: gap of unknown length
* 1689: contig of 1294 bp in length
* 1690: gap of unknown length
* 1691: contig of 1294 bp in length
* 1692: gap of unknown length
* 1693: contig of 1294 bp in length
* 1694: gap of unknown length
* 1695: contig of 1294 bp in length
* 1696: gap of unknown length
* 1697: contig of 1294 bp in length
* 1698: gap of unknown length
* 1699: contig of 1294 bp in length
* 1700: gap of unknown length
* 1701: contig of 1294 bp in length
* 1702: gap of unknown length
* 1703: contig of 1294 bp in length
* 1704: gap of unknown length
* 1705: contig of 1294 bp in length
* 1706: gap of unknown length
* 1707: contig of 1294 bp in length
* 1708: gap of unknown length
* 1709: contig of 1294 bp in length
* 1710: gap of unknown length
* 1711: contig of 1294 bp in length
* 1712: gap of unknown length
* 1713: contig of 1294 bp in length
* 1714: gap of unknown length
* 1715: contig of 1294 bp in length
* 1716: gap of unknown length
* 1717: contig of 1294 bp in length
* 1718: gap of unknown length
* 1719: contig of 1294 bp in length
* 1720: gap of unknown length
* 1721: contig of 1294 bp in length
* 1722: gap of unknown length
* 1723: contig of 1294 bp in length
* 1724: gap of unknown length
* 1725: contig of 1294 bp in length
* 1726: gap of unknown length
* 1727: contig of 1294 bp in length
* 1728: gap of unknown length
* 1729: contig of 1294 bp in length
* 1730: gap of unknown length
* 1731: contig of 1294 bp in length
* 1732: gap of unknown length
* 1733: contig of 1294 bp in length
* 1734: gap of unknown length
* 1735: contig of 1294 bp in length
* 1736: gap of unknown length
* 1737: contig of 1294 bp in length
* 1738: gap of unknown length
* 1739: contig of 1294 bp in length
* 1740: gap of unknown length
* 1741: contig of 1294 bp in length
* 1742: gap of unknown length
* 1743: contig of 1294 bp in length
* 1744: gap of unknown length
* 1745: contig of 1294 bp in length
* 1746: gap of unknown length
* 1747: contig of 1294 bp in length
* 1748: gap of unknown length
* 1749: contig of 1294 bp in length
* 1750: gap of unknown length
* 1751: contig of 1294 bp in length
* 1752: gap of unknown length
* 1753: contig of 1294 bp in length
* 1754: gap of unknown length
* 1755: contig of 1294 bp in length
* 1756: gap of unknown length
* 1757: contig of 1294 bp in length
* 1758: gap of unknown length
* 1759: contig of 1294 bp in length
* 1760: gap of unknown length
* 1761: contig of 1294 bp in length
* 1762: gap of unknown length
* 1763: contig of 1294 bp in length
* 1764: gap of unknown length
* 1765: contig of 1294 bp in length
* 1766: gap of unknown length
* 1767: contig of 1294 bp in length
* 1768: gap of unknown length
* 1769: contig of 1294 bp in length
* 1770: gap of unknown length
* 1771: contig of 1294 bp in length
* 1772: gap of unknown length
* 1773: contig of 1294 bp in length
* 1774: gap of unknown length
* 1775: contig of 1294 bp in length
* 1776: gap of unknown length
* 1777: contig of 1294 bp in length
* 1778: gap of unknown length
* 1779: contig of 1294 bp in length
* 1780: gap of unknown length
* 1781: contig of 1294 bp in length
* 1782: gap of unknown length
* 1783: contig of 1294 bp in length
* 1784: gap of unknown length
* 1785: contig of 1294 bp in length
* 1786: gap of unknown length
* 1787: contig of 1294 bp in length
* 1788: gap of unknown length
* 1789: contig of 1294 bp in length
* 1790: gap of unknown length
* 1791: contig of 1294 bp in length
* 1792: gap of unknown length
* 1793: contig of 1294 bp in length
* 1794: gap of unknown length
* 1795: contig of 1294 bp in length
* 1796: gap of unknown length
* 1797: contig of 1294 bp in length
* 1798: gap of unknown length
* 1799: contig of 1294 bp in length
* 1800: gap of unknown length
* 1801: contig of 1294 bp in length
* 1802: gap of unknown length
* 1803: contig of 1294 bp in length
* 1804: gap of unknown length
* 1805: contig of 1294 bp in length
* 1806: gap of unknown length
* 1807: contig of 1294 bp in length
* 1808: gap of unknown length
* 1809: contig of 1294 bp in length
* 1810: gap of unknown length
* 1811: contig of 1294 bp in length
* 1812: gap of unknown length
* 1813: contig of 1294 bp in length
* 1814: gap of unknown length
* 1815: contig of 1294 bp in length
* 1816: gap of unknown length
* 1817: contig of 1294 bp in length
* 1818: gap of unknown length
* 1819: contig of 1294 bp in length
* 1820: gap of unknown length
* 1821: contig of 1294 bp in length
* 1822: gap of unknown length
* 1823: contig of 1294 bp in length
* 1824: gap of unknown length
* 1825: contig of 1294 bp in length
* 1826: gap of unknown length
* 1827: contig of 1294 bp in length
* 1828: gap of unknown length
* 1829: contig of 1294 bp in length
* 1830: gap of unknown length
* 1831: contig of 1294 bp in length
* 1832: gap of unknown length
* 1833: contig of 1294 bp in length
* 1834: gap of unknown length
* 1835: contig of 1294 bp in length
* 1836: gap of unknown length
* 1837: contig of 1294 bp in length
* 1838: gap of unknown length
* 1839: contig of 1294 bp in length
* 1840: gap of unknown length
* 1841: contig of 1294 bp in length
* 1842: gap of unknown length
* 1843: contig of 1294 bp in length
* 1844: gap of unknown length
* 1845: contig of 1294 bp in length
* 1846: gap of unknown length
* 1847: contig of 1294 bp in length
* 1848: gap of unknown length
* 1849: contig of 1294 bp in length
* 1850: gap of unknown length
* 1851: contig of 1294 bp in length
* 1852: gap of unknown length
* 1853: contig of 1294 bp in length
* 1854: gap of unknown length
* 1855: contig of 1294 bp in length
* 1856: gap of unknown length
* 1857: contig of 1294 bp in length
* 1858: gap of unknown length
* 1859: contig of 1294 bp in length
* 1860: gap of unknown length
* 1861: contig of 1294 bp in length
* 1862: gap of unknown length
* 1863: contig of 1294 bp in length
* 1864: gap of unknown length
* 1865: contig of 1294 bp in length
* 1866: gap of unknown length
* 1867: contig of 1294 bp in length
* 1868: gap of unknown length
* 1869: contig of 1294 bp in length
* 1870: gap of unknown length
* 1871: contig of 1294 bp in length
* 1872: gap of unknown length
* 1873: contig of 1294 bp in length
* 1874: gap of unknown length
* 1875: contig of 1294 bp in length
* 1876: gap of unknown length
* 1877: contig of 1294 bp in length
* 1878: gap of unknown length
* 1879: contig of 1294 bp in length
* 1880: gap of unknown length
* 1881: contig of 1294 bp in length
* 1882: gap of unknown length
* 1883: contig of 1294 bp in length
* 1884: gap of unknown length
* 1885: contig of 1294 bp in length
* 1886: gap of unknown length
* 1887: contig of 1294 bp in length
* 1888: gap of unknown length
* 1889: contig of 1294 bp in length
* 1890: gap of unknown length
* 1891: contig of 1294 bp in length
* 1892: gap of unknown length
* 1893: contig of 1294 bp in length
* 1894: gap of unknown length
* 1895: contig of 1294 bp in length
* 1896: gap of unknown length
* 1897: contig of 1294 bp in length
* 1898: gap of unknown length
* 1899: contig of 1294 bp in length
* 1900: gap of unknown length
* 1901: contig of 1294 bp in length
* 1902: gap of unknown length
* 1903: contig of 1294 bp in length
* 1904: gap of unknown length
* 1905: contig of 1294 bp in length
* 1906: gap of unknown length
* 1907: contig of 1294 bp in length
* 1908: gap of unknown length
* 1909: contig of 1294 bp in length
* 1910: gap of unknown length
* 1911: contig of 1294 bp in length
* 1912: gap of unknown length
* 1913: contig of 1294 bp in length
* 1914: gap of unknown length
* 1915: contig of 1294 bp in length
* 1916: gap of unknown length
* 1917: contig of 1294 bp in length
* 1918: gap of unknown length
* 1919: contig of 1294 bp in length
* 1920: gap of unknown length
* 1921: contig of 1294 bp in length
* 1922: gap of unknown length
* 1923: contig of 1294 bp in length
* 1924: gap of unknown length
* 1925: contig of 1294 bp in length
* 1926: gap of unknown length
* 1927: contig of 1294 bp in length
* 1928: gap of unknown length
* 1929: contig of 1294 bp in length
* 1930: gap of unknown length
* 1931: contig of 1294 bp in length
* 1932: gap of unknown length
* 1933: contig of 1294 bp in length
* 1934: gap of unknown length
* 1935: contig of 1294 bp in length
* 1936: gap of unknown length
* 1937: contig of 1294 bp in length
* 1938: gap of unknown length
* 1939: contig of 1294 bp in length
* 1940: gap of unknown length
* 1941: contig of 1294 bp in length
* 1942: gap of unknown length
* 1943: contig of 1294 bp in length
* 1944: gap of unknown length
* 1945: contig of 1294 bp in length
* 1946: gap of unknown length
* 1947: contig of 1294 bp in length
* 1948: gap of unknown length
* 1949: contig of 1294 bp in length
* 1950: gap of unknown length
* 1951: contig of 1294 bp in length
* 1952: gap of unknown length
* 1953: contig of 1294 bp in length
* 1954: gap of unknown length
* 1955: contig of 1294 bp in length
* 1956: gap of unknown length
* 1957: contig of 1294 bp in length
* 1958: gap of unknown length
* 1959: contig of 1294 bp in length
* 1960: gap of unknown length
* 1961: contig of 1294 bp in length
* 1962: gap of unknown length
* 1963: contig of 1294 bp in length
* 1964: gap of unknown length
* 1965: contig of 1294 bp in length
* 1966: gap of unknown length
* 1967: contig of 1294 bp in length
* 1968: gap of unknown length
* 1969: contig of 1294 bp in length
* 1970: gap of unknown length
* 1971: contig of 1294 bp in length
* 1972: gap of unknown length
* 1973: contig of 1294 bp in length
* 1974: gap of unknown length
* 1975: contig of 1294 bp in length
* 1976: gap of unknown length
* 1977: contig of 1294 bp in length
* 1978: gap of unknown length
* 1979: contig of 1294 bp in length
* 1980: gap of unknown length
* 1981: contig of 1294 bp in length
* 1982: gap of unknown length
* 1983: contig of 1294 bp in length
* 1984: gap of unknown length
* 1985: contig of 1294 bp in length
* 1986: gap of unknown length
* 1987: contig of 1294 bp in length
* 1988: gap of unknown length
* 1989: contig of 1294 bp in length
* 1990: gap of unknown length
* 1991: contig of 1294 bp in length
* 1992: gap of unknown length
* 1993: contig of 1294 bp in length
* 1994: gap of unknown length
* 1995: contig of 1294 bp in length
* 1996: gap of unknown length
* 1997: contig of 1294 bp in length
* 1998: gap of unknown length
* 1999: contig of 1294 bp in length
* 2000: gap of unknown length
* 2001: contig of 1294 bp in length
* 2002: gap of unknown length
* 2003: contig of 1294 bp in length
* 2004: gap of unknown length
* 2005: contig of 1294 bp in length
* 2006: gap of unknown length
* 2007: contig of 1294 bp in length
* 2008: gap of unknown length
* 2009: contig of 1294 bp in length
* 2010: gap of unknown length
* 2011: contig of 1294 bp in length
* 2012: gap of unknown length
* 2013: contig of 1294 bp in length
* 2014: gap of unknown length
* 2015: contig of 1294 bp in length
* 2016: gap of unknown length
* 2017: contig of 1294 bp in length
* 2018: gap of unknown length
* 2019: contig of 1294 bp in length
* 2020: gap of unknown length
* 2021: contig of 1294 bp in length
* 2022: gap of unknown length
* 2023: contig of 1294 bp in length
* 2024: gap of unknown length
* 2025: contig of 1294 bp in length
* 2026: gap of unknown length
* 2027: contig of 1294 bp in length
* 2028: gap of unknown length
* 2029: contig of 1294 bp in length
* 2030: gap of unknown length
* 2031: contig of 1294 bp in length
* 2032: gap of unknown length
* 2033: contig of 1294 bp in length
* 2034: gap of unknown length
* 2035: contig of 1294 bp in length
* 2036: gap of unknown length
* 2037: contig of 1294 bp in length
* 2038: gap of unknown length
* 2039: contig of 1294 bp in length
* 2040: gap of unknown length
* 2041: contig of 1294 bp in length
* 2042: gap of unknown length
* 2043: contig of 1294 bp in length
* 2044: gap of unknown length
* 2045: contig of 1294 bp in length
* 2046: gap of unknown length
* 2047: contig of 1294 bp in length
* 2048: gap of unknown length
* 2049: contig of 1294 bp in length
* 2050: gap of unknown length
* 2051: contig of 1294 bp in length
* 2052: gap of unknown length
* 2053: contig of 1294 bp in length
* 2054: gap of unknown length
* 2055: contig of 1294 bp in length
* 2056: gap of unknown length
* 2057: contig of 1294 bp in length
* 2058: gap of unknown length
* 2059: contig of 1294 bp in length
* 2060: gap of unknown length
* 2061: contig of 1294 bp in length
* 2062: gap of unknown length
* 2063: contig of 1294 bp in length
* 2064: gap of unknown length
* 2065: contig of 1294 bp in length
* 2066: gap of unknown length
* 2067: contig of 1294 bp in length
* 2068: gap of unknown length
* 2069: contig of 1294 bp in length
* 2070: gap of unknown length
* 2071: contig of 1294 bp in length
* 2072: gap of unknown length
* 2073: contig of 1294 bp in length
* 2074: gap of unknown length
* 2075: contig of 1294 bp in length
* 2076: gap of unknown length
* 2077: contig of 1294 bp in length
* 2078: gap of unknown length
* 2079: contig of 1294 bp in length
* 2080: gap of unknown length
* 2081: contig of 1294 bp in length
* 2082: gap of unknown length
* 2083: contig of 1294 bp in length
* 2084: gap of unknown length
* 2085: contig of 1294 bp in length
* 2086: gap of unknown length
* 2087: contig of 1294 bp in length
* 2088: gap of unknown length
* 2089: contig of 1294 bp in length
* 2090: gap of unknown length
* 2091: contig of 1294 bp in length
* 2092: gap of unknown length
* 2093: contig of 1294 bp in length
* 2094: gap of unknown length
* 2095: contig of 1294 bp in length
* 2096: gap of unknown length
* 2097: contig of 1294 bp in length
* 2098: gap of unknown length
* 2099: contig of 1294 bp in length
* 2100: gap of unknown length
* 2101: contig of 1294 bp in length
* 2102: gap of unknown length
* 2103: contig of 1294 bp in length
* 2104: gap of unknown length
* 2105: contig of 1294 bp in length
* 2106: gap of unknown length
* 2107: contig of 1294 bp in length
* 2108: gap of unknown length
* 2109: contig of 1294 bp in length
* 2110: gap of unknown length
* 2111: contig of 1294 bp in length
* 2112: gap of unknown length
* 2113: contig of 1294 bp in length
* 2114: gap of unknown length
* 2115: contig of 1294 bp in length
* 2116: gap of unknown length
* 2117: contig of 1294 bp in length
*

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 30, 2004, 15:31:34 ; Search time 1840 Seconds
(without alignments)
18574.323 Million cell updates/sec

Title: US-09-913-878A-1
Perfect score: 8045
Sequence: 1 ggtaccggccccctcgca.....cgtattacaatcccgatcac 8045

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_23Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 8045 | 100.0 | 8045 | 3 | AA65171 Neurospor |
| 2 | 220.8 | 2.7 | 5163 | 2 | AAV20700 Cryptospor |
| 3 | 220.8 | 2.7 | 5163 | 3 | AA61849 ORF encod |
| 4 | 220.8 | 2.7 | 5163 | 6 | ABT04778 C parvum |
| 5 | 220.8 | 2.7 | 5318 | 3 | AA61848 DNA encod |
| 6 | 220.8 | 2.7 | 5318 | 6 | ABT04777 C parvum |
| 7 | 219.2 | 2.7 | 5318 | 2 | AAV20701 Cryptospor |
| 8 | 211.4 | 2.6 | 1236 | 6 | AA44410 Human lun |
| 9 | 205 | 2.5 | 567 | 3 | AA229550 HIV codon |
| 10 | 203.8 | 2.5 | 2336 | 4 | ABL25662 Drosophil |
| 11 | 200.4 | 2.5 | 5511 | 3 | AA61847 Cryptospor |
| 12 | 200.4 | 2.5 | 5511 | 6 | ABT04776 C parvum |
| 13 | 200.4 | 2.5 | 7334 | 3 | AA61846 Cryptospor |
| 14 | 200.4 | 2.5 | 7334 | 6 | ABT04775 C parvum |
| 15 | 196.6 | 2.4 | 486 | 4 | AA75507 Polyglut |
| 16 | 194.6 | 2.4 | 867 | 2 | AA91461 T. gondii |
| 17 | 194.6 | 2.4 | 867 | 2 | AA91460 T. gondii |
| 18 | 194.6 | 2.4 | 867 | 4 | AA542783 Reverse c |
| 19 | 194.6 | 2.4 | 867 | 4 | AA542784 Reverse c |
| 20 | 194.6 | 2.4 | 1397 | 2 | AA91462 T. gondii |
| 21 | 194.6 | 2.4 | 1397 | 2 | AA91463 T. gondii |
| 22 | 194.6 | 2.4 | 1397 | 4 | AA542786 Reverse c |
| 23 | 194.6 | 2.4 | 1397 | 4 | AA542785 T. gondii |

| | | | | | |
|----|-------|-----|--------|---|--------------------|
| 24 | 182.8 | 2.3 | 13631 | 4 | ABL20354 Drosophil |
| c | 177 | 2.2 | 7758 | 6 | ABL33103 Human imm |
| 25 | 176.6 | 2.2 | 198 | 2 | AAT17204 DNA-spann |
| 26 | 176.6 | 2.2 | 198 | 2 | AAT17205 DNA-spann |
| 27 | 176.6 | 2.2 | 198 | 2 | AAT17202 DNA-spann |
| 28 | 175 | 2.2 | 198 | 2 | AAT17203 DNA-spann |
| 29 | 175 | 2.2 | 198 | 2 | AAT17203 DNA-spann |
| c | 171.8 | 2.1 | 1400 | 5 | AA83195 DNA encod |
| 30 | 162.8 | 2.0 | 4829 | 4 | ABL13399 Drosophil |
| 31 | 162.8 | 2.0 | 4829 | 4 | ABL13399 Drosophil |
| c | 162.8 | 2.0 | 39746 | 6 | ABL13398 Drosophil |
| 32 | 160 | 2.0 | 2717 | 6 | ABQ76406 S. cerevi |
| 33 | 158.8 | 2.0 | 2000 | 7 | ACC61172 Gene sequ |
| 34 | 158.8 | 2.0 | 3218 | 6 | ABQ76319 S. cerevi |
| 35 | 158.8 | 2.0 | 3218 | 6 | ABQ76319 S. cerevi |
| 36 | 157.4 | 2.0 | 1972 | 9 | ADC87176 Human GPC |
| 37 | 157.4 | 2.0 | 2032 | 2 | AAT10117 Drosophil |
| 38 | 157.4 | 2.0 | 2032 | 4 | ABL05381 Drosophil |
| 39 | 157.4 | 2.0 | 2032 | 6 | AB53745 cDNA enco |
| 40 | 157.4 | 2.0 | 2032 | 6 | AB53745 cDNA enco |
| c | 157.4 | 2.0 | 2032 | 6 | AB53745 cDNA enco |
| 41 | 157.4 | 2.0 | 4282 | 4 | ABL05380 Drosophil |
| c | 157.4 | 2.0 | 6668 | 6 | ABL33697 Human imm |
| 42 | 156.8 | 1.9 | 6948 | 4 | AAH62808 Shrimp wh |
| 43 | 156.8 | 1.9 | 305107 | 4 | AAH62689 Shrimp wh |
| 44 | 156.8 | 1.9 | 305107 | 4 | AAH62689 Shrimp wh |
| 45 | 156.4 | 1.9 | 2550 | 4 | AA823437 Candida a |

ALIGNMENTS

RESULT 1
AA65171
ID AA65171 standard; DNA; 8045 BP.
XX AC AA65171;
AC AA65171;
DT 16-NOV-2000 (first entry)
DE Neurospora crassa qde-1 gene.
KW Gene silencing; quelling deficient; qde-1; al-1; ds.
XX OS Neurospora crassa.
XX PH Key Location/Qualifiers
FT CDS 2447..6655
FT /*tag= a
FT /product= "QDE-1"
XX WO2000050581-A2.
XX PD 31-AUG-2000.
XX PF 16-FEB-2000; 2000WO-IT000048.
XX PR 22-FEB-1999; 99IT-RM000117.
XX PA (UYRO-) UNIV ROMA LA SAPIENZA.
XX PI Macino G, Cogoni C;
XX DR WFI; 2000-579171/54.
XX DR P-PSDB; AAB13956.
XX PT Novel polynucleotide encoding a polypeptide which has a silencing
XX activity and comprising a RNA-dependent RNA polymerase domain.
XX PS Claim 1; Page 31-43; 48pp; English.
XX CC The present sequence is the Neurospora crassa qde-1 gene. This gene has
XX silencing activity. The qde-1 gene was isolated by mutational analysis of
XX an al-1 transgenic strain. This strain had an albino phenotype resulting
XX from post-transcriptional silencing of the endogenous al-1 gene.
XX Reversion of this phenotype indicated a mutation in a silencing gene. The
XX silencing gene, qde-1, could then be isolated. Modulation of qde-1
XX expression may be used to inactivate genes and to silence suppression of

| CC | genes | Db |
|----|--|----|
| XX | Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 U; 0 Other; | |
| SQ | Query Match 100.0%; Score 8045; DB 3; Length 8045; | |
| | Best Local Similarity 100.0%; Pred. No. 0; | |
| | Matches 8045; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |
| QY | 1 GGTACCGGCCCCCCCCCTCGAGGTGACAGCGGTATCGATAAGCTTGATCGAATTCGAAAA 60 | |
| Db | 1 GGTACCGGCCCCCCCCCTCGAGGTGACAGCGGTATCGATAAGCTTGATCGAATTCGAAAA 60 | |
| QY | 61 GTCTTAAAGTCTCTCTCTCATGCTAGGTAGGAGCCCTACTGCGTGTGTAGTGTAGGT 120 | |
| Db | 61 GTCTTAAAGTCTCTCTCTCATGCTAGGTAGGAGCCCTACTGCGTGTGTAGTGTAGGT 120 | |
| QY | 121 CCGAATAGGTTCGTACGCTGTCCTCTTAGGGGGCTGGTTTAAACAGGAGACAGGGTACA 180 | |
| Db | 121 CCGAATAGGTTCGTACGCTGTCCTCTTAGGGGGCTGGTTTAAACAGGAGACAGGGTACA 180 | |
| QY | 181 GTCTCTCGGAAACAGAGACGACCAAGCTTTGTGTGTGACATCATCTTTGAGACCATTCGT 240 | |
| Db | 181 GTCTCTCGGAAACAGAGACGACCAAGCTTTGTGTGTGACATCATCTTTGAGACCATTCGT 240 | |
| QY | 241 CTGAGGGTCCATACCAACCATACATCTGTGTAGAGAGTGGTGGGATATTAGAGTTACT 300 | |
| Db | 241 CTGAGGGTCCATACCAACCATACATCTGTGTAGAGAGTGGTGGGATATTAGAGTTACT 300 | |
| QY | 301 TTGAGAGTTTCTGAAGAGATTGAGACTGTAGGTACGTTTCATTTATCGGTGCGACGAGAT 360 | |
| Db | 301 TTGAGAGTTTCTGAAGAGATTGAGACTGTAGGTACGTTTCATTTATCGGTGCGACGAGAT 360 | |
| QY | 361 GGGCAGCATGAGTGCACATTTGCAATTTGTTGGTGGCAGCGACATCGATAGTATAC 420 | |
| Db | 361 GGGCAGCATGAGTGCACATTTGCAATTTGTTGGTGGCAGCGACATCGATAGTATAC 420 | |
| QY | 421 ATGTGTGGGCTCGAGACTGTCCAAATGGAACCAATGTGTAAGGCGGTGAGCGCGAGTGGGTC 480 | |
| Db | 421 ATGTGTGGGCTCGAGACTGTCCAAATGGAACCAATGTGTAAGGCGGTGAGCGCGAGTGGGTC 480 | |
| QY | 481 AACTTGAGTTATGTACAGTACTTGGGCTGTAGCGGCGAGCGGAGCATTCGGTCT 540 | |
| Db | 481 AACTTGAGTTATGTACAGTACTTGGGCTGTAGCGGCGAGCGGAGCATTCGGTCT 540 | |
| QY | 541 CGAATGACATTCCTCAAGCAGCAGCAACACTGTGTCCATTTAAAGCGTTTCTCCATGG 600 | |
| Db | 541 CGAATGACATTCCTCAAGCAGCAGCAACACTGTGTCCATTTAAAGCGTTTCTCCATGG 600 | |
| QY | 601 GGGAAACAATGGCGCGGTAGTATCTGTAAAGACCTTAGAGACCGGTATTAGAACTTTG 660 | |
| Db | 601 GGGAAACAATGGCGCGGTAGTATCTGTAAAGACCTTAGAGACCGGTATTAGAACTTTG 660 | |
| QY | 661 GGTGATCTATCTAGATCACCAGTGTGCCCTGATAGTCCATCAATAGTCTGGATGAGT 720 | |
| Db | 661 GGTGATCTATCTAGATCACCAGTGTGCCCTGATAGTCCATCAATAGTCTGGATGAGT 720 | |
| QY | 721 CCGAGATCAGTCCGGGATGTGTCACCGATATATCCATGACTTATCTCTGATTTATCCAG 780 | |
| Db | 721 CCGAGATCAGTCCGGGATGTGTCACCGATATATCCATGACTTATCTCTGATTTATCCAG 780 | |
| QY | 781 CGAGACCCCAATCAGTATGACCGACGACGACCATCCCGCGCGAGCGGTATGAGATCC 840 | |
| Db | 781 CGAGACCCCAATCAGTATGACCGACGACGACCATCCCGCGCGAGCGGTATGAGATCC 840 | |
| QY | 841 TGAACATATGTTAGGCATATGTATGGCATATGTATGATATGATAGCATGACCATATCTG 900 | |
| Db | 841 TGAACATATGTTAGGCATATGTATGGCATATGTATGATATGATAGCATGACCATATCTG 900 | |
| QY | 901 AGAGCTGCTGTGACTACTAGAGGTATGTAGAGTAAATGTACGTTACTTACTCTACC 960 | |
| Db | 901 AGAGCTGCTGTGACTACTAGAGGTATGTAGAGTAAATGTACGTTACTTACTCTACC 960 | |
| QY | 961 AAATGTATCCCGCAGTCACTCGTATTCCTCCGCTGTAGTGTACACTAGAACTAGCCCCGCT 1020 | |

| | | |
|------|---|------|
| 961 | AAATGTATCCCGCAGTCACTCGTATTCCTCCGCTGTAGTGTACACTAGCACTAGCCCCGCT | 1020 |
| 1021 | TTTGGATCTTGGAGTGTCCCATTCACGACCTTGCAGAAACCAAACTCGCAAAGCG | 1080 |
| 1021 | TTTGGATCTTGGAGTGTCCCATTCACGACCTTGCAGAAACCAAACTCGCAAAGCG | 1080 |
| 1081 | CGCTGACAGGACACCGTGTGCTGACCTTTGCGGCGCTTCTTCTTGTGACCGGGCCCA | 1140 |
| 1081 | CGCTGACAGGACACCGTGTGCTGACCTTTGCGGCGCTTCTTCTTGTGACCGGGCCCA | 1140 |
| 1141 | GGCAGACAGCCGCTGCTTCTTCTCAAGTGTACGAAGGCGATTAAAAATCCAAGA | 1200 |
| 1141 | GGCAGACAGCCGCTGCTTCTTCTCAAGTGTACGAAGGCGATTAAAAATCCAAGA | 1200 |
| 1201 | CAACAGATCCTGGAACCTTTGATCATTCCTCTGCTTACTCAGACGAAAAAATCTAAAAAC | 1260 |
| 1201 | CAACAGATCCTGGAACCTTTGATCATTCCTCTGCTTACTCAGACGAAAAAATCTAAAAAC | 1260 |
| 1261 | TTCAAGTCTGTTGTCAGATTAAATGAAACCGTTAAACAGGACATGTCGGCCCTTCTGCT | 1320 |
| 1261 | TTCAAGTCTGTTGTCAGATTAAATGAAACCGTTAAACAGGACATGTCGGCCCTTCTGCT | 1320 |
| 1321 | CTCATGCCCTCGAAAACCTCAGCTTCTTCCATCTCTCGTCCGATCCGCCCTTCCCAAGAC | 1380 |
| 1321 | CTCATGCCCTCGAAAACCTCAGCTTCTTCCATCTCTCGTCCGATCCGCCCTTCCCAAGAC | 1380 |
| 1381 | CTGATTCAGCAGCACTGACCTGTCTACTTTACTGTGGCTCTCTTAGAGTAGGGCTGG | 1440 |
| 1381 | CTGATTCAGCAGCACTGACCTGTCTACTTTACTGTGGCTCTCTTAGAGTAGGGCTGG | 1440 |
| 1441 | CCATCCCGGCGCCCTGCGAGCTTCTGCAAGTCAAAAGCGATGGAAGGATCGCGCTCCAGT | 1500 |
| 1441 | CCATCCCGGCGCCCTGCGAGCTTCTGCAAGTCAAAAGCGATGGAAGGATCGCGCTCCAGT | 1500 |
| 1501 | GGGTGCGGCATACAGCCATTCCAAATGTCTCTGTCGATGCGGAGAGAAAAATGGA | 1560 |
| 1501 | GGGTGCGGCATACAGCCATTCCAAATGTCTCTGTCGATGCGGAGAGAAAAATGGA | 1560 |
| 1561 | AGGCTAGAGCTCGGACACCTGTACGCTCACTCGGCTTACCCTTACCAACCACTGAGTGA | 1620 |
| 1561 | AGGCTAGAGCTCGGACACCTGTACGCTCACTCGGCTTACCCTTACCAACCACTGAGTGA | 1620 |
| 1621 | TATCATCAACCATACATCAGGCGCATACCGCGCGCTCTCTCGAATCATCGGCTGAAT | 1680 |
| 1621 | TATCATCAACCATACATCAGGCGCATACCGCGCGCTCTCTCGAATCATCGGCTGAAT | 1680 |
| 1681 | GTGACCAATTCGCGCGACAGCGGCTTTTTCGATCTCTGCGGAGCTTCTTCCACCGGTTT | 1740 |
| 1681 | GTGACCAATTCGCGCGACAGCGGCTTTTTCGATCTCTGCGGAGCTTCTTCCACCGGTTT | 1740 |
| 1741 | CGCTGTATCTGGGAGCGTATCGAGCTGGAATGGGACTGTTTACCACTCGAATCACTTGAC | 1800 |
| 1741 | CGCTGTATCTGGGAGCGTATCGAGCTGGAATGGGACTGTTTACCACTCGAATCACTTGAC | 1800 |
| 1801 | GGCTGGCGGATTCATGGCCATCCACCGGATGATCGTCCCTTCACTGTTTCTCTGC | 1860 |
| 1801 | GGCTGGCGGATTCATGGCCATCCACCGGATGATCGTCCCTTCACTGTTTCTCTGC | 1860 |
| 1861 | AACCTACTCTCCCGACAGATCAGACGAGGCTTTATGCCAGTTTCGGGCGTTGCCAATT | 1920 |
| 1861 | AACCTACTCTCCCGACAGATCAGACGAGGCTTTATGCCAGTTTCGGGCGTTGCCAATT | 1920 |
| 1921 | ATAGCCACAGCAGAGCTCGCAAGTTCGTTGCAATGCTCAAACTGTTTCAACACTT | 1980 |
| 1921 | ATAGCCACAGCAGAGCTCGCAAGTTCGTTGCAATGCTCAAACTGTTTCAACACTT | 1980 |
| 1981 | TGCAGAACAGGACCAACCGAAAAACCACTTGGCCGCTCGGCGGAAACCGGTGTCTT | 2040 |
| 1981 | TGCAGAACAGGACCAACCGAAAAACCACTTGGCCGCTCGGCGGAAACCGGTGTCTT | 2040 |
| 2041 | GAACTGATAGCCCGCAATCCCAACAGGTGCGCTTCCCGCTGCGCTGCTTGTCCC | 2100 |

Db 2041 GAACTCGATCAGCCCAATCCAAACAGGTCGCTTTCCCGCTGCGCTGTGTCC 2100
 Qy 2101 GACTTCGTGAGCAACAGACTGGCGCTATGGCCAGGCCCCATTCATGACAAATGGT 2160
 Db 2101 GACTTCGTGAGCAACAGACTGGCGCTATGGCCAGGCCCCATTCATGACAAATGGT 2160
 Qy 2161 TAAATCGAGTCAAAAATCTACAAGAGATCAAGCGCCGCAAGAGATATACCTGTGTCACCCC 2220
 Db 2161 TAAATCGAGTCAAAAATCTACAAGAGATCAAGCGCCGCAAGAGATATACCTGTGTCACCCC 2220
 Qy 2221 GACAACCGAAGCCCTGAAGGGCAAGCCACAGTGGTGTGTTTACTTCGGACACTTA 2280
 Db 2221 GACAACCGAAGCCCTGAAGGGCAAGCCACAGTGGTGTGTTTACTTCGGACACTTA 2280
 Qy 2281 TCGACCCCTTACGCCCAAGACGCTCCCGCTTGGTAGAAGCTCGAGTCGGGTAAATCAGGTAT 2340
 Db 2281 TCGACCCCTTACGCCCAAGACGCTCCCGCTTGGTAGAAGCTCGAGTCGGGTAAATCAGGTAT 2340
 Qy 2341 ATCTCACAGTGGCTTGTCTCTTTTCGGTGGTGGTGGCTTTTATCATGCTTTATCGCT 2400
 Db 2341 ATCTCACAGTGGCTTGTCTCTTTTCGGTGGTGGTGGCTTTTATCATGCTTTATCGCT 2400
 Qy 2401 AACCCGGTGTATCTTACTTCCAGGCCGATCTCCACGGCACAAATGAACCCCTATTAC 2460
 Db 2401 AACCCGGTGTATCTTACTTCCAGGCCGATCTCCACGGCACAAATGAACCCCTATTAC 2460
 Qy 2461 TCTTAGGAAGAGGAATAGCCCGCTCGAGGAATCATAAACCGGCTCAATACGACTCAA 2520
 Db 2461 TCTTAGGAAGAGGAATAGCCCGCTCGAGGAATCATAAACCGGCTCAATACGACTCAA 2520
 Qy 2521 CCTGGGCTTCAGTGTGCGAGACACAACTCTCACCCCCCAGCCGCGGAGGAGCTGGC 2580
 Db 2521 CCTGGGCTTCAGTGTGCGAGACACAACTCTCACCCCCCAGCCGCGGAGGAGCTGGC 2580
 Qy 2581 CGAGAGTGAAGAGGATTCGGTGGCCATGACAAGATCTACAGAGCCCTGAACTTCTCTA 2640
 Db 2581 CGAGAGTGAAGAGGATTCGGTGGCCATGACAAGATCTACAGAGCCCTGAACTTCTCTA 2640
 Qy 2641 CTGGCGGAGAGTACTCCCTGAACAGCGAGAGCAACTCTTCTCATCGAGGCCAAGC 2700
 Db 2641 CTGGCGGAGAGTACTCCCTGAACAGCGAGAGCAACTCTTCTCATCGAGGCCAAGC 2700
 Qy 2701 TCGAGCTCGAACTGGGTGCCCAAGCCACGCGACCCCTGACAGCTTCGGTGGTCCAA 2760
 Db 2701 TCGAGCTCGAACTGGGTGCCCAAGCCACGCGACCCCTGACAGCTTCGGTGGTCCAA 2760
 Qy 2761 GGAACCTCCCGCGCGCTACTGCGCGGCAACAATGGGCATTCGAGACTGTGTGCTCGA 2820
 Db 2761 GGAACCTCCCGCGCGCTACTGCGCGGCAACAATGGGCATTCGAGACTGTGTGCTCGA 2820
 Qy 2821 GGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACAGGTGAAAGTTTGGCAGAAC 2880
 Db 2821 GGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACAGGTGAAAGTTTGGCAGAAC 2880
 Qy 2881 TCTAAGCGGCCCCAAGTGGCTGAGCGGCCCAACCTCTACCAACCAAGCAAGGATGA 2940
 Db 2881 TCTAAGCGGCCCCAAGTGGCTGAGCGGCCCAACCTCTACCAACCAAGCAAGGATGA 2940
 Qy 2941 GCCCGCCAAATGTCATTTGCTGTATCCGCCCAACCGCTGTGACTGCTCTGCCACAGG 3000
 Db 2941 GCCCGCCAAATGTCATTTGCTGTATCCGCCCAACCGCTGTGACTGCTCTGCCACAGG 3000
 Qy 3001 TCTCTCTATTACCGCGCGGCGATACCCCTTAAAGTTCCCGATCCAGTGAATACCGGTT 3060
 Db 3001 TCTCTCTATTACCGCGCGGCGATACCCCTTAAAGTTCCCGATCCAGTGAATACCGGTT 3060
 Qy 3061 CAAACGACCATCTCTCGAGAGTGAATCTCAATCAGTGCACCAAGCGGCGCAAGGCA 3120
 Db 3061 CAAACGACCATCTCTCGAGAGTGAATCTCAATCAGTGCACCAAGCGGCGCAAGGCA 3120
 Qy 3121 GCTGTCTGATAATGTTCGCTGCGCGCGCCCGCTGCTATTTGAGCGCTTTTGA 3180
 Db 3121 GCTGTCTGATAATGTTCGCTGCGCGCGCCCGCTGCTATTTGAGCGCTTTTGA 3180

Qy 3181 CAAAGTACCGACTCGAAGCATGCCAATACGAGAGATCCACGGCACAGGTCATAGACG 3240
 Db 3181 CAAAGTACCGACTCGAAGCATGCCAATACGAGAGATCCACGGCACAGGTCATAGACG 3240
 Qy 3241 AGCGGACCGAGTGGATTCCTTTGATACATCTCAAGGCACTTCTATGTTGAGTGTCTT 3300
 Db 3241 AGCGGACCGAGTGGATTCCTTTGATACATCTCAAGGCACTTCTATGTTGAGTGTCTT 3300
 Qy 3301 CAGCGCTTCGGTCAATCAGAGCACTACCCAGAGTAGTTTTTGGGCTCTCTTCACA 3360
 Db 3301 CAGCGCTTCGGTCAATCAGAGCACTACCCAGAGTAGTTTTTGGGCTCTCTTCACA 3360
 Qy 3361 GCGCAGAGAAAGCGCTGTGATGCCACGCTTTTGAAGTGGACATTTGATTGAGTC 3420
 Db 3361 GCGCAGAGAAAGCGCTGTGATGCCACGCTTTTGAAGTGGACATTTGATTGAGTC 3420
 Qy 3421 TCTTAGCAAGGAAGAACCAACCAAGTCCACATAGATACCAAGCCCTTTCTCGTTC 3480
 Db 3421 TCTTAGCAAGGAAGAACCAACCAAGTCCACATAGATACCAAGCCCTTTCTCGTTC 3480
 Qy 3481 CCAGGTTGAAACTTCGTTTCAAGCTTACTATGAGTCTGTTTCCAAGTTCGGGCGGAGG 3540
 Db 3481 CCAGGTTGAAACTTCGTTTCAAGCTTACTATGAGTCTGTTTCCAAGTTCGGGCGGAGG 3540
 Qy 3541 CGCAATTCGGAGCGAGTGGCTCAAAATGACTGGCTCGAGCGAAGAAAGGCTGATC 3600
 Db 3541 CGCAATTCGGAGCGAGTGGCTCAAAATGACTGGCTCGAGCGAAGAAAGGCTGATC 3600
 Qy 3601 TCAGGTTCAAGTTCATGCCCCGCTGGTTCAGCTCGGCTGAGAAATATTTGGCCGAAAT 3660
 Db 3601 TCAGGTTCAAGTTCATGCCCCGCTGGTTCAGCTCGGCTGAGAAATATTTGGCCGAAAT 3660
 Qy 3661 TCCCAATGGCTACAGAAAGCTCCTCGCTGTCATGCGGAGTTTACAGACTCTTTAT 3720
 Db 3661 TCCCAATGGCTACAGAAAGCTCCTCGCTGTCATGCGGAGTTTACAGACTCTTTAT 3720
 Qy 3721 GCCTCAAAAGTAGACTTCGAAAGCAGAGCTGGGCTTAAAGTACGACCCCTTCCTGGTC 3780
 Db 3721 GCCTCAAAAGTAGACTTCGAAAGCAGAGCTGGGCTTAAAGTACGACCCCTTCCTGGTC 3780
 Qy 3781 TACCGCGCGGATGTCAGAGATATCTGAAGACTCTCTACCGGCTTGATGCTTCGTTG 3840
 Db 3781 TACCGCGCGGATGTCAGAGATATCTGAAGACTCTCTACCGGCTTGATGCTTCGTTG 3840
 Qy 3841 TAAACCTTTTCAGAAAGCCGCCAAGCTGTTGTCGACGCAATGACGGGCAACTT 3900
 Db 3841 TAAACCTTTTCAGAAAGCCGCCAAGCTGTTGTCGACGCAATGACGGGCAACTT 3900
 Qy 3901 TGAGAGCAAGGTAAGTGGCTTGTCTCTCTGTTCTAGACTACAAATCGGACAACTC 3960
 Db 3901 TGAGAGCAAGGTAAGTGGCTTGTCTCTCTGTTCTAGACTACAAATCGGACAACTC 3960
 Qy 3961 GCCTACTGCGCCCTTACTTGTGAAGCTGAAGCGCTCATGTTTCGAGCGGCTGTCG 4020
 Db 3961 GCCTACTGCGCCCTTACTTGTGAAGCTGAAGCGCTCATGTTTCGAGCGGCTGTCG 4020
 Qy 4021 ACTCACCCGTCGGTTCGGTTCCTGATAGGTTTTTTCGAGATCTTATACCGCTACGAG 4080
 Db 4021 ACTCACCCGTCGGTTCGGTTCCTGATAGGTTTTTTCGAGATCTTATACCGCTACGAG 4080
 Qy 4081 CACGAGCCAGTGTACCGCGTGTGACAAACCAAGCTGCGGTGAGAGAGTAT 4140
 Db 4081 CACGAGCCAGTGTACCGCGTGTGACAAACCAAGCTGCGGTGAGAGAGTAT 4140
 Qy 4141 CCAAGTGGCTCACGATGGGCAACATTTCTCTGTAGGCGCCCAATGGCGGCTTTCTTCG 4200
 Db 4141 CCAAGTGGCTCACGATGGGCAACATTTCTCTGTAGGCGCCCAATGGCGGCTTTCTTCG 4200
 Qy 4201 CAAAGATGCGGATACAGGAACCTCTCAGGAGGTTCCAGCTCCGCGCCGAGACCCGAA 4260
 Db 4201 CAAAGATGCGGATACAGGAACCTCTCAGGAGGTTCCAGCTCCGCGCCGAGACCCGAA 4260

| | | | |
|----|------|--|------|
| QY | 4261 | ACCATCATCAAGGAGAGAGTCCACTTCTTTTCCGAGACCGGCAATTACGTTCCGACCTGA | 4320 |
| Db | 4261 | ACCATCATCAAGGAGAGAGTCCACTTCTTTTCCGAGACCGGCAATTACGTTCCGACCTGA | 4320 |
| QY | 4321 | TGTGTTTCAAGACAGATCTGTGCTTCCGCGACAGAGAACCTGTAGAGCAACGACCGAGTT | 4380 |
| Db | 4321 | TGTGTTTCAAGACAGATCTGTGCTTCCGCGACAGAGAACCTGTAGAGCAACGACCGAGTT | 4380 |
| QY | 4381 | CAAAGTTAGTCAAATGCTGGACTGGCTCTGCAACTCGACAACAACACATTGGCAGCCGA | 4440 |
| Db | 4381 | CAAAGTTAGTCAAATGCTGGACTGGCTCTGCAACTCGACAACAACACATTGGCAGCCGA | 4440 |
| QY | 4441 | CCTCAAGTTGTTCTCCCGTATCCAGTCCGGTCTGAGTAAAGATATGCAATATGACATT | 4500 |
| Db | 4441 | CCTCAAGTTGTTCTCCCGTATCCAGTCCGGTCTGAGTAAAGATATGCAATATGACATT | 4500 |
| QY | 4501 | GGAGCTCACCAGATCAGACACCAAGACCGATCTTCTTTCACTTTCAGGCACTGGCGA | 4560 |
| Db | 4501 | GGAGCTCACCAGATCAGACACCAAGACCGATCTTCTTTCACTTTCAGGCACTGGCGA | 4560 |
| QY | 4561 | AGTGATGAATGACGGGTAGGCCGATGTCCGAAGCGTGGCCAGAGGATACGCGATGT | 4620 |
| Db | 4561 | AGTGATGAATGACGGGTAGGCCGATGTCCGAAGCGTGGCCAGAGGATACGCGATGT | 4620 |
| QY | 4621 | TCCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGCGGTTTGGTTGCGCCAAAGGAAT | 4680 |
| Db | 4621 | TCCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGCGGTTTGGTTGCGCCAAAGGAAT | 4680 |
| QY | 4681 | GTGGTTATTCGACGTTGACGACACAGCGATGAGGATTTGATTCGAGACATACCCGTCCCA | 4740 |
| Db | 4681 | GTGGTTATTCGACGTTGACGACACAGCGATGAGGATTTGATTCGAGACATACCCGTCCCA | 4740 |
| QY | 4741 | GGCAAGTGGGAATCGACATTCGTTGATAAACATCAACGTACCCCTCGAAGTCGGAGCGT | 4800 |
| Db | 4741 | GGCAAGTGGGAATCGACATTCGTTGATAAACATCAACGTACCCCTCGAAGTCGGAGCGT | 4800 |
| QY | 4801 | GGCTTCGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTGTTACCTGTCTCTGGAAGATAG | 4860 |
| Db | 4801 | GGCTTCGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTGTTACCTGTCTCTGGAAGATAG | 4860 |
| QY | 4861 | AGCAGGACAAGGTGAAGATGCGCAGGCAATCGGTACCGTCTTATCAAGCATTTGCA | 4920 |
| Db | 4861 | AGCAGGACAAGGTGAAGATGCGCAGGCAATCGGTACCGTCTTATCAAGCATTTGCA | 4920 |
| QY | 4921 | ACGACAGTTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAAGTGGAAATTTGCCAATGGGT | 4980 |
| Db | 4921 | ACGACAGTTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAAGTGGAAATTTGCCAATGGGT | 4980 |
| QY | 4981 | TTACGAGATTATTCAGTCCGGCACTCAGTTCAGCCAGCGCGGTGCGCTTTTCTTGC | 5040 |
| Db | 4981 | TTACGAGATTATTCAGTCCGGCACTCAGTTCAGCCAGCGCGGTGCGCTTTTCTTGC | 5040 |
| QY | 5041 | TGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAACAGTGGGTTCGATCC | 5100 |
| Db | 5041 | TGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAACAGTGGGTTCGATCC | 5100 |
| QY | 5101 | CAAGAGCAAAAGTACTTGTGAAGACATCGCGCTGGGATCTTCAAAAGCGGAAATGTGACAC | 5160 |
| Db | 5101 | CAAGAGCAAAAGTACTTGTGAAGACATCGCGCTGGGATCTTCAAAAGCGGAAATGTGACAC | 5160 |
| QY | 5161 | GTTCAAGTCAAGCTGAACATCCGTGTCGCTCGATCAGCATACATTTACATGATTCGCCA | 5220 |
| Db | 5161 | GTTCAAGTCAAGCTGAACATCCGTGTCGCTCGATCAGCATACATTTACATGATTCGCCA | 5220 |
| QY | 5221 | TTTCTGGGTTGCTTGAGGAAATAGGTTTCATGTCGGATTCTCTCAAAGTTCAGAGA | 5280 |
| Db | 5221 | TTTCTGGGTTGCTTGAGGAAATAGGTTTCATGTCGGATTCTCTCAAAGTTCAGAGA | 5280 |
| QY | 5281 | CGAGGAGAGTCTTTTACATCTCTATCGGACTGTGATGTCCTCGTGGCGGATCCCGAGC | 5340 |
| Db | 5281 | CGAGGAGAGTCTTTTACATCTCTATCGGACTGTGATGTCCTCGTGGCGGATCCCGAGC | 5340 |
| QY | 5341 | CCATTTCCCTAGTCATATCCAAACGGGTTTCGAGCAGTCTTCAAGCCAGAGTCCACAGTCT | 5400 |

| | | | |
|----|------|--|------|
| Db | 5341 | CCATTTCCCTTAGTGAATCAACGGGTTTGAGCAGTCTTCAAGCCAGAGCTCCACAGTCT | 5400 |
| Qy | 5401 | CAAGGATGTAATCATCTTCTCTACTAAAGGAGATGACCGCTTGTCTAAGAGCTATCTCG | 5460 |
| Db | 5401 | CAAGGATGTAATCATCTTCTCTACTAAAGGAGATGACCGCTTGTCTAAGAGCTATCTCG | 5460 |
| Qy | 5461 | TGGAGACTACGCGGATATGCGCTTGGCTGCTCGGATCCGGAGATCGTGATGGTTT | 5520 |
| Db | 5461 | TGGAGACTACGCGGCGATATGGCCCTGGTCTGCTCGGATCCGAGATCGTGATGGTTT | 5520 |
| Qy | 5521 | CGTCAATGCGGAATGCCTCTGGAGCCGACCTGTCTAGTACCTTAAGAAGACAAAAC | 5580 |
| Db | 5521 | CGTCAATGCGGAATGCCTCTGAGCCGACCTGTCTAGTACCTTAAGAAGACAAAAC | 5580 |
| Qy | 5581 | GACTTTCAACAACACTTATGCGCTCAACGGCAGCGGCTCAGCGGCCAAAGACAGACTAC | 5640 |
| Db | 5581 | GACTTTCAACAACACTTATGCGCTCAACGGCAGCGGCTCAGCGGCCAAAGACAGACTAC | 5640 |
| Qy | 5641 | ATACGATATGATCCAGAAGAGCTTCCATTTGCGCCCTCGAGCCCAACTTCTTGGCATGTG | 5700 |
| Db | 5641 | ATACGATATGATCCAGAAGAGCTTCCATTTGCGCCCTCGAGCCCAACTTCTTGGCATGTG | 5700 |
| Qy | 5701 | CATTAACCTCAAAAGAAAGCGTCTGTTACATCAACAATAGTGTCTTAAACCGCGCCAT | 5760 |
| Db | 5701 | CATTAACCTCAAAAGAAAGCGTCTGTTACATCAACAATAGTGTCTTAAACCGCGCCAT | 5760 |
| Qy | 5761 | CATTCTTAGTTCACTGTTGGGAAACCTCGTCTGATCAGACCAAGAGTATTGCTTTAA | 5820 |
| Db | 5761 | CATTCTTAGTTCACTGTTGGGAAACCTCGTCTGATCAGACCAAGAGTATTGCTTTAA | 5820 |
| Qy | 5821 | CGAAGCAAGCTGGGCTCAATTCGTAGGAACTTGTCTTGGCGTGCATGTCCTTCTGTA | 5880 |
| Db | 5821 | CGAAGCAAGCTGGGCTCAATTCGTAGGAACTTGTCTTGGCGTGCATGTCCTTCTGTA | 5880 |
| Qy | 5881 | CCCNACTCAAGAGCGACAGTTGGCTCGGGCGGAGAGCCTTACCACATTTAGACTA | 5940 |
| Db | 5881 | CCCNACTCAAGAGCGACAGTTGGCTCGGGCGGAGAGCCTTACCACATTTAGACTA | 5940 |
| Qy | 5941 | CCTGAAATCTCCATCGCCAGGCGCTCGATTTGAACAGAACTCGAAGCCTTCCACATGC | 6000 |
| Db | 5941 | CCTGAAATCTCCATCGCCAGGCGCTCGATTTGAACAGAACTCGAAGCCTTCCACATGC | 6000 |
| Qy | 6001 | CATGAAAGCGGCCAAGATACAGAACGCGCGTCACTTTTGGGATCCGGATCTCGCTTC | 6060 |
| Db | 6001 | CATGAAAGCGGCCAAGATACAGAACGCGCGTCACTTTTGGGATCCGGATCTCGCTTC | 6060 |
| Qy | 6061 | CTACTTACACGTTCTTCAAGAGATTAGCGACAAGTCCGATCGTCCGACTGCTATTAC | 6120 |
| Db | 6061 | CTACTTACACGTTCTTCAAGAGATTAGCGACAAGTCCGATCGTCCGACTGCTATTAC | 6120 |
| Qy | 6121 | GACTCTGAAGAACCGTATCGCGGAGTGCAGAAAGATATGCGCAGGTTGCTCAAAACAA | 6180 |
| Db | 6121 | GACTCTGAAGAACCGTATCGCGGAGTGCAGAAAGATATGCGCAGGTTGCTCAAAACAA | 6180 |
| Qy | 6181 | GGAGATGAGAGACAGCAAGGACCCCTACCTGTCGCGTCAACACAGGTTTATGAAAATG | 6240 |
| Db | 6181 | GGAGATGAGAGACAGCAAGGACCCCTACCTGTCGCGTCAACACAGGTTTATGAAAATG | 6240 |
| Qy | 6241 | GTGCGCCATCAACGCTGAGCGGATGGAACAATCCGGAGCAAAATATGATTCTAAGTGAT | 6300 |
| Db | 6241 | GTGCGCCATCAACGCTGAGCGGATGGAACAATCCGGAGCAAAATATGATTCTAAGTGAT | 6300 |
| Qy | 6301 | CAGGTTCGTGAGCTGTCCTTCTCGCGGACCGTGAGTGAATACATGCGGATGCTGAG | 6360 |
| Db | 6301 | CAGGTTCGTGAGCTGTCCTTCTCGCGGACCGTGAGTGAATACATGCGGATGCTGAG | 6360 |
| Qy | 6361 | GGCTAGCACGGCTTCAAGCTGTACTACCAAGAGCCCCCAAGTTCGTGTCGAGATGGC | 6420 |
| Db | 6361 | GGCTAGCACGGCTTCAAGCTGTACTACCAAGAGCCCCCAAGTTCGTGTCGAGATGGC | 6420 |
| Qy | 6421 | GGGCAGACAGCTCCGCTACATTANGCGCGAGTGAACGAGCAGCCCGTCAAGCGCCCC | 6480 |

6421 GGGCAGACAGCTCGGTACATTAGCGCAGATGACGAGACGCCGGTGAAGCGCCCC 6480
6481 GCGTTGATGACCGGTTCACTGATGCGGGCTTGATGCGGGATAGAGTTTACGAGCA 6540
6481 GCGCTGTGATGACCGGTTCACTGATGCGGGCTTGATGCGGGATAGAGTTTACGAGCA 6540
6541 GTATGTGCGCAGCTGAGGGGCGATGCGAGTACCTGATCGGAGGTCTATGAAGT 6600
6541 GTATGTGCGCAGCTGAGGGGCGATGCGAGTACCTGATCGGAGGTCTATGAAGT 6600
6601 GCTGGCGATGATGATTTTGTGATGGAATGGTTTACAGGGAATGCGGATTTAGTCGGG 6660
6601 GCTGGCGATGATGATTTTGTGATGGAATGGTTTACAGGGAATGCGGATTTAGTCGGG 6660
6661 TTGCTTTGAGGGGGCGGTCATCTTATTTTCTGATTTTCCGAGTCTTTGGTCTGGG 6720
6661 TTGCTTTGAGGGGGCGGTCATCTTATTTTCTGATTTTCCGAGTCTTTGGTCTGGG 6720
6721 AATGGGAATCTTCTGACCATGTTCCCTATGGGTGAGTCTCATTTTATGGACTTCAGTTT 6780
6721 AATGGGAATCTTCTGACCATGTTCCCTATGGGTGAGTCTCATTTTATGGACTTCAGTTT 6780
6781 TGGAGTTTCTTTCGGCTTGGATCTCCCTGAGCAAGCGGGTGGGACATTCGGTTG 6840
6781 TGGAGTTTCTTTCGGCTTGGATCTCCCTGAGCAAGCGGGTGGGACATTCGGTTG 6840
6841 GTCCGTAGAACATAAAAGAAATGGGAGATGATATGATTTCTTCGTCGCGTGATAATTT 6900
6841 GTCCGTAGAACATAAAAGAAATGGGAGATGATATGATTTCTTCGTCGCGTGATAATTT 6900
6901 GAGATATTTGGTGTCTGTTGATTCATGTCGAGATTTCCCTTCTTCTACTTGTGGCC 6960
6901 GAGATATTTGGTGTCTGTTGATTCATGTCGAGATTTCCCTTCTTCTACTTGTGGCC 6960
6961 CGGAGAGCAGATTGGTGCACATTTCAACAATAGTCAGCAACCTTTGAGGAACTATCGA 7020
6961 CGGAGAGCAGATTGGTGCACATTTCAACAATAGTCAGCAACCTTTGAGGAACTATCGA 7020
7021 TCTCGATACAGCGTTAGATTAATAGCGTCTCCCTTCTAGCTCCGACCTCCATC 7080
7021 TCTCGATACAGCGTTAGATTAATAGCGTCTCCCTTCTAGCTCCGACCTCCATC 7080
7081 CTAGGTGAGACCCCGGATCTCTAGACTCTTTTCAGACCTGTCTAAGTGAGATATATA 7140
7081 CTAGGTGAGACCCCGGATCTCTAGACTCTTTTCAGACCTGTCTAAGTGAGATATATA 7140
7141 GTGCAAGCAGTTGGTGTGTTACACTTTGGTATTTACATCAACAGACAATTTGAGATGACT 7200
7141 GTGCAAGCAGTTGGTGTGTTACACTTTGGTATTTACATCAACAGACAATTTGAGATGACT 7200
7201 CATTCCTCGGTTTCTTCGATTCGACTCTATTTGATGATTTTTCATCTACCC 7260
7201 CATTCCTCGGTTTCTTCGATTCGACTCTATTTGATGATTTTTCATCTACCC 7260
7261 CTGACAGCATACCTCTAGCAAACTTTTGTGTAATGATAGGCTCTTTGGGATTC 7320
7261 CTGACAGCATACCTCTAGCAAACTTTTGTGTAATGATAGGCTCTTTGGGATTC 7320
7321 GAGTTTGGCTTGAACCTTTGATCTTTTGTCTTTGCTTTTCTTTGCTTTGCTTTGAAA 7380
7321 GAGTTTGGCTTGAACCTTTGATCTTTTGTCTTTTGTCTTTTCTTTGCTTTGCTTTGAAA 7380
7381 GTATATCAACTCTACCTGATCACTCGGAGCAACAACAACAACAACAACAACAACAACA 7440
7381 GTATATCAACTCTACCTGATCACTCGGAGCAACAACAACAACAACAACAACAACAACA 7440
7441 ACA 7500
7441 ACA 7500
7501 ACA 7560
7501 ACA 7560

7561 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7620
7561 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7620
7621 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7680
7621 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7680
7681 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7740
7681 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7740
7741 ACCGAGCGACCTCATAGCTTTATGGGAAAATAATACACGTCGAGATACATACAAATAC 7800
7741 ACCGAGCGACCTCATAGCTTTATGGGAAAATAATACACGTCGAGATACATACAAATAC 7800
7801 AGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7860
7801 AGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7860
7861 CACCACCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACACATGAACCGTCGAG 7920
7861 CACCACCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACACATGAACCGTCGAG 7920
7921 ATCCCTATAGTGTGATGATGATGCGGCGCGAATTCGATATCAAGCTTATCGATACCGT 7980
7921 ATCCCTATAGTGTGATGATGATGCGGCGCGAATTCGATATCAAGCTTATCGATACCGT 7980
7981 CGACCTCGAGGGGGGGCGCGGATACCCCAATTCGCGCTATAGTGTGATGATGATGATGAT 8040
7981 CGACCTCGAGGGGGGGCGCGGATACCCCAATTCGCGCTATAGTGTGATGATGATGATGAT 8040
8041 GTCAC 8045
8041 GTCAC 8045

RESULT 2
AAV20700
ID AAV20700 standard; DNA; 5163 BP.
XX AC
XX AC AAV20700;
XX 17-AUG-1998 (first entry)
XX Cryptosporidium parvum GP900 antigen open reading frame.
DE Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
XX antibody; prophylaxis; treatment; inhibition; retardation; detection;
KW diagnosis; human; ds.
XX
XX Cryptosporidium parvum.
OS
XX Key Location/Qualifiers
FH 1. 5163
CDS /*tag= a
FT /product= "GP900 antigen"
FT misc_feature 524..1270
FT /*tag= b
FT /note= "region containing NINC mutations"
XX
XX WO9806430-A1.
FN XX
XX 19-FEB-1998.
PD XX
XX 11-AUG-1997; 97WO-US014104.
PF XX
XX 14-AUG-1996; 96US-00700651.
PR XX
XX (REGC) UNIV CALIFORNIA.
PA Petersen C, Leesch J, Nelson RC, Gut J;
PI

[illegible]

| | |
|----------|---|
| RESULT 3 | |
| AAA61849 | |
| ID | AAA61849 standard; DNA; 5163 BP. |
| XX | |
| AC | AAA61849; |
| XX | |
| DT | 28-OCT-2000 (first entry) |
| XX | |
| DE | ORF encoding a portion of Cryptosporidium parvum NINC isolate GP900. |
| XX | |
| KW | GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis; |
| KW | competitive inhibition; attachment; invasion; ligand binding; sporozoite; |
| KW | merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds. |
| XX | |
| OS | Cryptosporidium parvum. |

| | | |
|----|---|---------------------|
| XX | Key | Location/Qualifiers |
| PH | 1569. .7182 | |
| FT | /tag= a | |
| FT | /partial | |
| FT | /product= "Cryptosporidium parvum NINC isolate GP900" | |
| FT | /note= "No start or stop codons given in the | |
| FT | specification" | |
| FT | | |
| XX | US6071518-A. | |
| PN | | |
| XX | 06-JUN-2000. | |
| PD | | |
| XX | 13-SEP-1997; | 97US-00928361. |
| XX | | |
| XX | 25-MAY-1992; | 92US-00891301. |
| PR | 01-JUN-1993; | 93US-00071880. |
| PR | 03-APR-1995; | 95US-00415751. |
| PR | 14-AUG-1996; | 96US-00700651. |
| PR | 13-SEP-1996; | 96US-0026062P. |
| PR | | |
| XX | (REGC) UNIV CALIFORNIA. | |
| PA | | |
| XX | Petersen C; | |
| XX | | |
| XX | WPI: 2000-422065/36. | |
| DR | P-PSDB; AAB11727. | |
| DR | | |
| XX | New GP900 protein fragments and fusion proteins of Cryptosporidium | |
| FT | parvum, useful for detecting the presence of the parasite and diagnosing | |
| FT | or treating Cryptosporidium infections by competitive inhibition of the | |
| FT | function of GP900. | |
| XX | | |
| XX | Claim 16; Col 47-52; 59pp; English. | |
| PS | | |
| XX | | |
| XX | The invention relates to the GP900 glycoprotein of the protozoan | |
| CC | Cryptosporidium parvum. DNA encoding it, GP900 fragments, and fusion | |
| CC | proteins comprising GP900 fragments. The invention also relates to the | |
| CC | administration of GP900 or fragments thereof to a host to elicit anti- | |
| CC | GP900 antibody production, and to a method of cryptosporidiosis treatment | |
| CC | or prophylaxis comprising administration of anti-GP900 antibodies to an | |
| CC | individual. Cryptosporidium parvum GP900 and GP900 fragments are able to | |
| CC | competitively inhibit sporozoite or merozoite attachment or invasion, and | |
| CC | are also useful for the generation of anti-GP900 antibodies. The | |
| CC | antibodies also inhibit sporozoite or merozoite attachment/invasion and | |
| CC | additionally inhibit the binding of GP900 ligands to GP900. GP900 | |
| CC | proteins, fragments and antibodies may therefore be used to treat or | |
| CC | prevent cryptosporidiosis. Infection with Cryptosporidium is a common | |
| CC | cause of diarrhoea in humans and causes life-threatening diarrhoea in | |
| CC | immunocompromised persons. Cryptosporidiosis can be contracted from | |
| CC | contaminated municipal water supplies (e.g., public swimming pools). It | |
| CC | is also a cause of disease in animals, resulting in financial losses in | |
| CC | agriculture. GP900 fragments, fusion proteins and antibodies may also be | |
| CC | used for the diagnosis of Cryptosporidium parvum infections, and for the | |
| CC | detection of the parasite in the environment. The present sequence | |
| CC | represents the open reading frame (ORF) encoding a portion of the GP900 | |
| CC | protein of the NINC isolate of Cryptosporidium parvum | |
| XX | | |
| XX | Sequence 5163 BP: 1873 A: 1138 C: 875 G: 1277 T: 0 U: 0 Other: | |

[illegible]

[illegible]

RESULT 4
ABT04778

AB104778
ID ABT04778 standard; DNA; 5163 bp.

AC ABT04778;

DT 27-SEP-2002 (first entry)

DE C parvum GP900 gene fragment SEQ ID NO: 4.

XX Cryptosporidium detection; GP900; P68; cryptosporidiosis;
KW gene; ds.

AA
OS
Cryptosporidium parvum.

PN WO200194631-A1.

PD 13-DEC-2001.

14-MAY-2001: 2001WO-US015624.

PR 06-JUN-2000; 2000US-00588995.

AA
PA (REGC) UNIV CALIFORNIA.

PI Petersen C, Barnes DA, Nelson RC, Gut J;

WPI; 2002-566447/60.

AA Detecting *Cryptosporidium* in biological and environmental samples and
PT diagnosis of cryptosporidiosis involves, contacting the sample with
PT *Cryptosporidium* G900. P68 or cryptopain antigen, antibody, DNA or RNA.

XX
PS
Disclosure: Page 103-104; 157pp; English:

xx The present invention relates to a method of detecting *Cryptosporidium* in
cc biological and environmental samples, and of diagnosing
cc cryptosporidiosis. This involves obtaining a sample and contacting it
cc with *Cryptosporidium* GP200, P68 or cryptosporin antigen, antibody, DNA or
cc RNA, or its variant, mutant or fragment. The method is also useful for
cc detecting and identifying individual *Cryptosporidium* isolates based on
cc the genetic characteristics, and for diagnosis of prior or concurrent
cc *Cryptosporidium* infection. The present sequence is a *C. parvum* coding
cc sequence used in the exemplification of the invention

| Seq | Sequence | 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 U; 0 Other; |
|-----------------------|-----------------|--|
| Query Match | 2.7%; | Score 220.8; DB 6; Length 5163; |
| Best Local Similarity | 67.0%; | Pred. No. 2.8e-36; |
| Matches 343; | Conservative 0; | Mismatches 167; Indels 2; Gaps 1 |
| Qy | 7402 | ACATCGGAACGACACACAAACAAACGACGACACACACACACACACACAAACAA 7461 |
| Db | 667 | ACTACGACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 726 |
| Qy | 7462 | ACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7521 |
| Db | 727 | ACAAACAGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 786 |
| Qy | 7522 | ACAACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7591 |
| Db | 787 | ACAACTAACACCAACTACTACAAACAAACAACTACTAACCAACAAACAAAC 846 |
| Qy | 7592 | ACAACACACACACACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7641 |
| Db | 847 | ACTACACACCAACTACTACACCAACAAACAACTTACAACCAACAAACCAAC 906 |
| Qy | 7642 | ACAACACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7701 |
| Db | 907 | ACTACCGAGAAACGACAAACAAACAACTACAAACAAACAAACAAACAAAC 965 |
| Qy | 7702 | GACAACATACAGGTTTAACTCTAGTACCAATTTGTAAACAGACCGAGCGAC 7761 |
| Db | 966 | TACAACCCACCAACAAACAAACAAACAAACAACTACAACCTACCAAGAAAC 1025 |
| Qy | 7762 | TGGGAATAATATACAGTCCAGAGATACATTAACAAATACAGCAACAAATGA 7821 |
| Db | 1026 | TACTACTACTACCAACAAACAAACAACTACTTACTTACCACAAACAAACAA 1084 |
| Qy | 7822 | AGAGTTTCTCGCGCGCGTTCACCAACCAACCAACCAACCAACCAACCAAC 7881 |
| Db | 1085 | CTACTACTACAAACAAACAAACAAACAAACAAACAAACAAACAACTACC 1144 |
| Qy | 7882 | CGAGTGTATGAGTGACGAGTACAAATGAACC 7913 |
| Db | 1145 | CAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1176 |

RESULT 5

AAA61848

ID AAA61848 standard; DNA; 5318 BP.

AC
AAA61848:

28-OCT-2000 (first entry)

XX DNA encoding a portion of *Cryptosporidium parvum* NINC isolate GP900.

| | |
|----|---|
| XX | GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis; |
| KW | competitive inhibition; attachment; invasion; ligand binding; sporozoite; |
| KW | merozoite; diarrhoea; protozoacids; ds. |

XX OS Cryptosporidium parvum.

| | | |
|-----------|------------|----------------------------|
| AX | | Location/Qualifiers |
| FH | Key | |
| FT | CDS | 1669 .7182 |

```

3'UTR
Ft      /partial
Ft      /*cay= a
Ft      /product= "Cryptosporidium parvum NINC isolate GP900"
Ft      /note= "No start codon given in the specification"
Ft      5167. 5318
Ft      /*tag= b

```

XX PN US6071518-A.

XX
PD 06-JUN-2000.XX
PF 12-SEP-1997: 97US-00928361.

[illegible]

| | |
|----|---|
| KW | Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis; |
| KW | gene; db. |
| XX | |
| OS | Cryptosporidium parvum. |
| XX | |
| PN | WO200194631-A1. |
| XX | |
| XX | 13-DEC-2001. |
| XX | |
| XX | 14-MAY-2001; 2001WO-US015624. |
| PF | |
| XX | |
| XX | 06-JUN-2000; 2000US-00588995. |
| PR | |
| XX | (REGC) UNIV CALIFORNIA. |
| PA | |
| XX | Petersen C, Barnes DA, Nelson RG, Gut J; |
| PI | |
| XX | WPI; 2002-566447/60. |
| DR | |
| XX | |
| XX | Detecting Cryptosporidium in biological and environmental samples and |
| PT | diagnosis of cryptosporidiosis involves, contacting the sample with |
| PT | Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA. |
| PT | |
| XX | |
| XX | Disclosure; Page 99-101; 157pp; English. |
| PS | |
| XX | |
| XX | The present invention relates to a method of detecting Cryptosporidium in |
| CC | biological and environmental samples, and of diagnosing |
| CC | cryptosporidiosis. This involves obtaining a sample and contacting it |
| CC | with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or |
| CC | RNA, or its variant, mutant or fragment. The method is also useful for |
| CC | detecting and identifying individual Cryptosporidium isolates based on |
| CC | the genetic characteristics, and for diagnosis of prior or concurrent |
| CC | Cryptosporidium infection. The present sequence is a C. parvum coding |
| CC | sequence used in the exemplification of the invention |
| XX | |
| XX | Sequence 5311 BP; 1964 A; 1204 C; 944 G; 1399 T; 0 U; 0 Other; |
| SO | |

[illegible]

[illegible][illegible]

Search completed: March 30, 2004, 17:07:30
Job time : 1845 secs

[illegible]

RESULT 4
US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6013882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

[illegible][illegible]

RESULT 5
 US-08-928-361B-3
 ; Sequence 3, Application US/08928361B
 ; Patent No. 6071518
 ; GENERAL INFORMATION:
 ; APPLICANT: Petersen, Carolyn
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
 ; TITLE OF INVENTION: SPECIES INFECTIONS
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
 ; STREET: 385 Sherman Avenue, Suite 6
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306-1840
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,361B
 ; FILING DATE: 12-SEP-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026,062
 ; FILING DATE: 13-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vetry, Hana
 ; REGISTRATION NUMBER: 30,518
 ; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-324-1677
 ; TELEFAX: 650-324-1678
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5318 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-928-361B-3

[illegible]

[illegible]


```

, , REFERENCE/DOCKET NUMBER: 480.76-1(HV)
, , TELECOMMUNICATION INFORMATION:
, , TELEPHONE: 650-324-1677
, , TELEFAX: 650-324-1678
, , INFORMATION FOR SEQ ID NO: 1:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 7334 base pairs
, , TYPE: nucleic acid
, , STRANDEDNESS: double
, , TOPOLOGY: linear
, , MOLECULE TYPE: DNA (genomic)
, , US-08-928-361B-1

```

[illegible]

```

RESULT 10
US-09-588-995A-1
, Sequence 1, Application US/09588995A
, Patent No. 6514697
, GENERAL INFORMATION:
, APPLICANT: PETERSEN, CAROLYN
, APPLICANT: BARNES, DEBBA A.
, APPLICANT: NELSON, RICHARD C.
, APPLICANT: GUT, JIRI
, TITLE OF INVENTION: METHODS FOR DETECTING
, TITLE OF INVENTION: ISOLATES AND FORMULATIONS
, TITLE OF INVENTION: INFECTIONS
, FILE REFERENCE: 480.19-5
, CURRENT APPLICATION NUMBER: US/09/588-995A
, CURRENT FILING DATE: 2000-06-06
, PRIOR APPLICATION NUMBER: 08/927,177
, PRIOR FILING DATE: 1997-03-27
, PRIOR APPLICATION NUMBER: 08/928,361
, PRIOR FILING DATE: 1997-09-12
, PRIOR APPLICATION NUMBER: 08/700,653
, PRIOR FILING DATE: 1996-08-14
, PRIOR APPLICATION NUMBER: 08/415,753
, PRIOR FILING DATE: 1995-04-03

```

[illegible]

```

RESULT 11
US-09-216-393B-340
; Sequence 340, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393B
; CURRENT FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/594,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 340
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
; OTHER INFORMATION:
US-09-216-393B-340

```

Query Match 2.4%; Score 194.6; DB 4; Length 867;
Best Local Similarity 79.6%; Pred. No. 2e-38;
Matches 230: Conservative 0; Mismatches 59; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 16:28:24 ; Search time 12100 Seconds
(without alignments)
19854.648 Million cell updates/sec

Title: US-09-913-878A-1
Perfect score: 8045
Sequence: 1 ggtaccgggccccccctcga.....cgtattacaatcccgatcac 8045

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *

1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 275 | 3.4 | 869 | 14 | CK159167 FGAS04056 |
| C 2 | 263 | 3.3 | 477 | 14 | CA743784 wr1s.pko |
| C 3 | 255.2 | 3.2 | 603 | 12 | BJ366946 BJ366946 |
| C 4 | 254.2 | 3.2 | 356 | 12 | BJ337970 BJ337970 |

RESULT 1
CK159167/c

LOCUS CK159167
DEFINITION FGAS040564 Triticum aestivum FGAS: Talt5 Triticum aestivum cDNA,
mRNA sequence.

ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 869)
Allard, P., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,
Peniket, C., Roach, J.L. and Sarhan, F.
Functional Genomics of Abiotic Stress in Wheat and Carola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033

ALIGNMENTS

| 5 | 254 | 3.2 | 494 | 12 | BJ360881 |
|----|-------|-----|------|----|----------|
| 6 | 250 | 3.1 | 394 | 9 | AU060224 |
| 7 | 247.4 | 3.1 | 1933 | 11 | AY325173 |
| 8 | 245.8 | 3.1 | 969 | 11 | AY383694 |
| 9 | 242.6 | 3.0 | 359 | 12 | BJ366856 |
| 10 | 242.6 | 3.0 | 627 | 12 | BJ426316 |
| 11 | 241.6 | 3.0 | 450 | 12 | BJ371297 |
| 12 | 238.8 | 3.0 | 889 | 14 | CK159613 |
| 13 | 235.8 | 2.9 | 436 | 14 | CB791539 |
| 14 | 234.8 | 2.9 | 256 | 14 | CA702278 |
| 15 | 234.8 | 2.9 | 732 | 28 | BZ284533 |
| 16 | 234.8 | 2.9 | 503 | 12 | BM656118 |
| 17 | 228.6 | 2.8 | 711 | 28 | BZ098560 |
| 18 | 227 | 2.8 | 906 | 29 | CNS03GJN |
| 19 | 225.8 | 2.8 | 827 | 29 | CNS04156 |
| 20 | 225.4 | 2.8 | 494 | 12 | EM641696 |
| 21 | 225.4 | 2.8 | 546 | 13 | C22974 |
| 22 | 225.4 | 2.8 | 681 | 29 | CNS02EOD |
| 23 | 224.4 | 2.8 | 608 | 12 | BJ330059 |
| 24 | 223.2 | 2.8 | 551 | 12 | BJ366220 |
| 25 | 223.2 | 2.8 | 571 | 9 | AU284550 |
| 26 | 222.8 | 2.8 | 612 | 10 | BE195101 |
| 27 | 221.8 | 2.8 | 970 | 29 | CNS03H6V |
| 28 | 219.8 | 2.7 | 904 | 29 | CNS03H6V |
| 29 | 218.8 | 2.7 | 818 | 28 | BZ229437 |
| 30 | 218.6 | 2.7 | 563 | 12 | BJ371247 |
| 31 | 218.2 | 2.7 | 652 | 13 | BM243685 |
| 32 | 216.4 | 2.7 | 429 | 9 | AU052930 |
| 33 | 215 | 2.7 | 617 | 9 | AU033655 |
| 34 | 214.2 | 2.7 | 280 | 12 | BJ377836 |
| 35 | 214.2 | 2.7 | 745 | 13 | BQ986791 |
| 36 | 214.2 | 2.7 | 854 | 13 | EX078177 |
| 37 | 213.8 | 2.7 | 576 | 14 | CD096608 |
| 38 | 213 | 2.6 | 878 | 28 | BH153470 |
| 39 | 212.8 | 2.6 | 410 | 12 | BJ370460 |
| 40 | 212.8 | 2.6 | 652 | 12 | BF000526 |
| 41 | 212.2 | 2.6 | 795 | 29 | CNS03HRI |
| 42 | 211.8 | 2.6 | 317 | 12 | BJ366180 |
| 43 | 211.8 | 2.6 | 419 | 9 | AU033391 |
| 44 | 211.8 | 2.6 | 544 | 9 | AU037837 |
| 45 | 211.2 | 2.6 | 303 | 12 | BJ367635 |

[illegible]

```
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc8k12"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Query Match      3.2%; Score 254; DB 12; Length 494;
Best Local Similarity 79.5%; Pred. No. 2.7e-21;
Matches 310; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 7341 TTGATCTTTTCTTTTGGCTTTTCGTTGAGTTGGAAAGTATATCACTCTACCTGAT 7400
Db 103 TTAATCTTTGATAGTTTACATTTCATGTTGTAACGCAAAATTCACCAACCAACAGAAC 162
QY 7401 CACATCGGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7460
Db 163 AAGAACAACAACCTACACCAACCAACCTACTACGACACACACACCAACCAACCAAC 222
QY 7461 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 7520
Db 223 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 282
QY 7521 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 7580
Db 283 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 342
QY 7581 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 7640
Db 343 AACANGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 402
QY 7641 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 7699
Db 403 AACACACACACACACACACACACACACACACACACACACACACACACACACACAC 462
QY 7700 AAGACAACCTATCATGCTTTAACTCAGTACC 7729
Db 463 AACANNANCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 492

RESULT 6
AU060224 394 bp mRNA linear EST 20-MAY-1999
DEFINITION Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
LOCUS dictyostelium cDNA clone SLA610, mRNA sequence.
ACCESSION AU060224.1 GI:4981328
VERSION AU060224.1
KEYWORDS Dictyostelium discoideum
SOURCE Dictyostelium discoideum
ORGANISM Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 394)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
TITLE Institute of Biological Sciences
JOURNAL University of Tsukuba
COMMENT 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hidekobiol.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES
source
1. .394
Location/Qualifiers
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"

/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc8k12"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Query Match      3.1%; Score 250; DB 9; Length 394;
Best Local Similarity 91.1%; Pred. No. 8.8e-21;
Matches 265; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 7402 ACATCGGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7461
Db 101 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 160
QY 7462 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7521
Db 161 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 220
QY 7522 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7581
Db 221 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 280
QY 7582 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7641
Db 281 ACCATCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 340
QY 7642 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7692
Db 341 ACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 391

RESULT 7
AY325173 1933 bp mRNA linear HTC 26-JUL-2003
LOCUS Rattus norvegicus Aa2-050 mRNA, complete cds.
DEFINITION Rattus norvegicus
ACCESSION AY325173
VERSION AY325173.1 GI:33086523
KEYWORDS HTC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1933)
AUTHORS Xu,C.S., Li,W.Q., Li,Y.C., Yang,K.J., Yan,H.M., Chang,C.F.,
Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,
Chai,L.Q., Yuan,J.Y., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
Liver regeneration after PH
Unpublished
JOURNAL 2 (bases 1 to 1933)
REFERENCE Xu,C.S., Li,W.Q., Li,Y.C., Yang,K.J., Yan,H.M., Chang,C.F.,
Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,
Chai,L.Q., Yuan,J.Y., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
Direct Submission
JOURNAL Submitted (16-JUN-2003) Henan Bioengineering Key Lab, Henan Normal
University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
China
FEATURES
source
1. .1933
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
41..1927
/note="liver regeneration-related protein LRRG146"
/codon_start=1
/product="Aa2-050"
/protein_id="AAP92574.1"
/db_xref="GI:33086524"
/translation="MAAGLDQGGPARCAGAGPPPLLRDSILAFVLTQMTGNEFKLN
QPDEGISVKEFSPNTSDFLLVSWDTSVRLVDVDPANSLRYQHTGAFLDCAFYVEN
LVGTHDAPRCVEYCPENVNMTVGTWSDPTKLVMDPTFCNAGTFSPQPKGYLVLSIEG
RVAVEYLDPSPEVQKKYAFKCHRLKENNIEOIYPVNAISLSQYFQHCHSNDGTTLA
IASSVYEMDDTEHPEDGTFPIQVTDATKPKSRKSSWNICGLSQDGHRLAASSWRL"

FEATURES
source
1. .1933
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
41..1927
/note="liver regeneration-related protein LRRG146"
/codon_start=1
/product="Aa2-050"
/protein_id="AAP92574.1"
/db_xref="GI:33086524"
/translation="MAAGLDQGGPARCAGAGPPPLLRDSILAFVLTQMTGNEFKLN
QPDEGISVKEFSPNTSDFLLVSWDTSVRLVDVDPANSLRYQHTGAFLDCAFYVEN
LVGTHDAPRCVEYCPENVNMTVGTWSDPTKLVMDPTFCNAGTFSPQPKGYLVLSIEG
RVAVEYLDPSPEVQKKYAFKCHRLKENNIEOIYPVNAISLSQYFQHCHSNDGTTLA
IASSVYEMDDTEHPEDGTFPIQVTDATKPKSRKSSWNICGLSQDGHRLAASSWRL"
```


[illegible]

| | | | | | | | | |
|------------|---|-------------------------------|-------------------|--------|-----------------|--|--|--|
| RESULT 12 | CK159613 | 889 bp | mrna | linear | EST 05-DEC-2003 | | | |
| LOCUS | FGAS041079 | Triticum aestivum FGAS: TaUT5 | Triticum aestivum | cdna, | | | | |
| DEFINITION | mRNA sequence. | | | | | | | |
| ACCESSION | CK159613 | | | | | | | |
| VERSION | CK159613.1 | GI:389885955 | | | | | | |
| KEYWORDS | EST. | | | | | | | |
| SOURCE | Triticum aestivum (bread wheat) | | | | | | | |
| ORGANISM | Triticum aestivum | | | | | | | |
| REFERENCE | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum. | | | | | | | |
| AUTHORS | 1 (bases 1 to 889) | | | | | | | |
| | Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Jarosche, A., Links, M.G., McCarthy, E.L., Monroy, A., Murak, I., Nilson, D., Pennak, C., Roach, J.L. and Sarhan, F. | | | | | | | |
| TITLE | Functional Genomics of Abiotic Stress In Wheat and Canola Crops | | | | | | | |
| JOURNAL | Unpublished (2003) | | | | | | | |
| COMMENT | Contact: Wm L Crosby | | | | | | | |
| | Bioinformatics | | | | | | | |
| | University of Saskatchewan, Department of Computer Science | | | | | | | |
| | 1C101 Engineering Building, 57 Campus Drive, Saskatoon, | | | | | | | |
| | Saskatchewan, S7N 5A9, Canada | | | | | | | |
| | Tel: 306 965 1769 | | | | | | | |
| | Fax: 306 965 2033 | | | | | | | |

```

Email: fgas_eefs@cs.usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [123,668].
Plate: Tal539_row: N column: 07.
      Location/Qualifiers
1..889
  /organism="Triticum aestivum"
  /mol_type="mRNA"
  /cultivar="Wheat line PI 178383"
  /db_xref="taxon:4565"
  /lab_host="DH5 alpha"
  /clone_lib="Triticum aestivum FGAS: Tal5"
  /notes="Organ: Crown; Vector: pGEM-T; SSH (suppression
subtractive hybridization) cDNA library from genotype

```

ORIGIN

```
Query Match          3.0%; Score 238.8; DB 14; Length 889;
Best Local Similarity 85.0%; Pred. No. 1.5e-19;
Matches 267; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

[illegible]

RESULT 13

| | | | | | |
|------------|-------------------------|----------------------------------|------|--------|-----------------|
| CB791539 | CE791539 | 436 bp | mRNA | linear | EST 16-MAY-2003 |
| LOCUS | AMGNMUC:NRH7-00028-F7-A | nrh7-00028-f7 5', mRNA sequence. | | | |
| DEFINITION | CB791539 | | | | |
| ACCESSION | CB791539 | | | | |
| VERSION | CB791539.1 | GI:29879932 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | Rattus norvegicus | (Norway rat) | | | |
| ORGANISM | Rattus norvegicus | | | | |

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Angen EST Program.
 Angen kat EST Program
 Unpublished (2003)
 Contact: Dan Fitzpatrick
 Angen, Inc
 One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
 Tel: 805 447-4881

FEATURES

```

1. 436
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrhy7-00028-f7"
/clone_lib="nrhy7 (10850)"
/note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; W Rat
hypothalamus adult female Wistar rat avg. Insert size 2.3
kb fraction 6 and 7"

```

ORIGIN

```
Query Match      2.9%; Score 235.8; DB 14; Length 436;
Best Local Similarity 87.5%; Pred. No. 4.3e-19;
Matches 258; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

7398 GATCAGTGGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7457
 136 GACAACAACGCAATGACACGACGACGACGACGACGACGACGACGACGAC 195
 7458 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7517
 196 AACCAACAACGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 255
 7518 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7577
 256 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 315
 7578 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7637
 316 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 375
 7638 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7692
 376 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 430

RESULT 14
 CA702278 256 bp mRNA linear EST 26-NOV-2002
 LOCUS wdk1c.pk005.a18 wdk1c Triticum aestivum cDNA clone wdk1c.pk005.a18
 DEFINITION 5' end, mRNA sequence.
 ACCESSION CA702278
 VERSION CA702278.1 GI:25424071
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 REFERENCE 1 (bases 1 to 256)
 AUTHORS Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
 Miao, G., Caraher, N., and Hanafey, M.K.
 TITLE DuPont Wheat cDNA Sequence
 JOURNAL Unpublished (2002)
 COMMENT Contact: Scott V. Tingey
 Crop Genetics
 E. I. DuPont de Nemours and Company
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
 Tel: 302-631-2602
 Fax: 302-631-2607
 Email: Scott.V.Tingey@USA.dupont.com
 Seq primer: M13.

FEATURES
 source
 1..256
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone="wdk1c.pk005.a18"
 /tissue_type="kernel"
 /clone_lib="wdk1c"
 /note="Vector: pBluescript SK-; Wheat (Triticum aestivum
 L.) developing kernel, 3 days after anthesis."

ORIGIN
 Query Match 2.9%; Score 234.8; DB 14; Length 256;
 Best Local Similarity 96.4%; Pred. No. 6.9e-19;
 Matches 239; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

7413 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7472
 1 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 60
 7473 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7532
 61 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 120
 7533 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7592

121 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 180
 7593 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7652
 181 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 240
 7653 AACCAACA 7660
 241 AACCAACA 248

RESULT 15
 BZ284533/c 732 bp DNA linear GSS 15-OCT-2002
 LOCUS CH230-385E7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
 DEFINITION CH230-385E7, genomic survey sequence.
 ACCESSION BZ284533
 VERSION BZ284533.1 GI:24014662
 KEYWORDS GSS.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 732)
 AUTHORS Zhao S., Shetty, J., Shatsman, S., Teegave, G., Geer, K.,
 Sivartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D.,
 Riggs, F., de Jong, P., and Fraser, C.M.
 Rat BAC End Sequences from Library CHORI-230 MboI segment
 Unpublished (1999)
 Other GSSs: CH230-385E7.TJC
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Fax: 301 838 0200
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/or_ering_information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 385 row: E column: 7
 Seq primer: T7
 Class: BAC ends.

FEATURES
 Location/Qualifiers
 1..732
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SENHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-385E7"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 2"
 /note="Vector: pTAREBAC1.3; Site 1: MboI; Site 2: MboI;
 CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by
 Pieter de Jong"

ORIGIN
 Query Match 2.9%; Score 234.8; DB 28; Length 732;
 Best Local Similarity 88.8%; Pred. No. 4.7e-19;
 Matches 254; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

7410 AGCAACAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7469
 374 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 315
 7470 AACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7529

```
Db      314 AGCAGCAACACAAACACAGCAACACAAACAAACAGCAGCAACAAACAAAC 255
QY      7530 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7589
Db      254 AGCAGCAGCAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 195
QY      7590 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7649
Db      194 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 135
QY      7650 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7695
Db      134 AGCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 89
```

Search completed: March 31, 2004, 02:00:38
Job time : 12105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:19:26 ; Search time 66 Seconds
(without alignments)
6001.994 Million cell updates/sec

Title: US-09-913-878A-2

Perfect score: 7397

Sequence: 1 MNPITPRKNSPVEIINRL.....YEVIGDDDFDGIQTNGNDY 1402

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Genesep1980a:*
- 2: Genesep1990a:*
- 3: Genesep2000a:*
- 4: Genesep2001a:*
- 5: Genesep2002a:*
- 6: Genesep2003a:*
- 7: Genesep2003bs:*
- 8: Genesep2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 7397 | 100.0 | 1402 | 3 | AAE13956 |
| 2 | 328 | 4.4 | 1196 | 4 | AAE67235 |
| 3 | 318 | 4.3 | 836 | 3 | AAE28532 |
| 4 | 297 | 4.0 | 1114 | 4 | AAU10006 |
| 5 | 297 | 4.0 | 1114 | 4 | AAE00897 |
| 6 | 292 | 3.9 | 1164 | 3 | AAE28831 |
| 7 | 280 | 3.8 | 1125 | 3 | AAE28833 |
| 8 | 267.5 | 3.6 | 1108 | 3 | AAE28530 |
| 9 | 236 | 3.2 | 417 | 3 | AAE28529 |
| 10 | 182 | 2.5 | 1219 | 3 | AAE38679 |
| 11 | 179.5 | 2.4 | 966 | 4 | AAU78795 |
| 12 | 176 | 2.4 | 1555 | 6 | ABJ25640 |
| 13 | 176 | 2.4 | 1832 | 6 | ABJ26240 |
| 14 | 173.5 | 2.3 | 1271 | 4 | ABE65816 |
| 15 | 168 | 2.2 | 1226 | 6 | ABE52986 |
| 16 | 165.5 | 2.2 | 2291 | 5 | AAU75351 |
| 17 | 164.5 | 2.2 | 4386 | 4 | ABG07375 |
| 18 | 163.5 | 2.2 | 4274 | 4 | ABG00972 |
| 19 | 163.5 | 2.2 | 4397 | 4 | ABG21944 |
| 20 | 160.5 | 2.2 | 1024 | 5 | ABE57200 |
| 21 | 159.5 | 2.2 | 1596 | 7 | AAE37929 |
| 22 | 158.5 | 2.1 | 1736 | 5 | AAU84308 |
| 23 | 157 | 2.1 | 1026 | 6 | ABO14710 |
| 24 | 156.5 | 2.1 | 1057 | 6 | ABO14709 |
| 25 | 156 | 2.1 | 1013 | 4 | AAU38678 |

| | | | | | |
|----|-------|-----|------|---|----------|
| 26 | 154 | 2.1 | 1025 | 4 | AAU38680 |
| 27 | 153.5 | 2.1 | 1828 | 4 | ABE50164 |
| 28 | 153 | 2.1 | 571 | 5 | ABE65975 |
| 29 | 133 | 2.1 | 2001 | 6 | ABU70405 |
| 30 | 133 | 2.1 | 5373 | 4 | AAU14603 |
| 31 | 153 | 2.1 | 5447 | 4 | AAU14897 |
| 32 | 152 | 2.1 | 1127 | 4 | ABE58310 |
| 33 | 152 | 2.1 | 2194 | 4 | AAU40114 |
| 34 | 151.5 | 2.0 | 218 | 4 | AAE00898 |
| 35 | 151 | 2.0 | 4873 | 6 | ABO14747 |
| 36 | 150 | 2.0 | 5909 | 4 | ABG23295 |
| 37 | 150 | 2.0 | 6819 | 4 | ABG23329 |
| 38 | 149 | 2.0 | 2462 | 6 | ABU61813 |
| 39 | 148.5 | 2.0 | 1418 | 6 | ABE53853 |
| 40 | 148.5 | 2.0 | 1457 | 3 | AAU93418 |
| 41 | 148.5 | 2.0 | 1605 | 4 | ABE70375 |
| 42 | 146 | 2.0 | 530 | 6 | ADA55193 |
| 43 | 146 | 2.0 | 1493 | 4 | ABE69228 |
| 44 | 145.5 | 2.0 | 1400 | 6 | ABE96237 |
| 45 | 145 | 2.0 | 1920 | 4 | ABE64441 |

ALIGNMENTS

RESULT 1
AAE13956
ID AAE13956 standard; protein; 1402 AA.
XX
AC AAE13956;
XX
AC
DT 16-NOV-2000 (first entry)
XX
DE Neurospora crassa QDE-1 protein.
XX
KW Gene silencing; quelling deficient; qde-1; al-1.
XX
OS Neurospora crassa.
XX
PN WO2000050581-A2.
XX
PD 31-AUG-2000.
XX
PF 16-FEB-2000; 2000WO-IT000048.
XX
PR 22-FEB-1999; 99IT-RM000117.
XX
PR (UYRO-) UNIV ROMA LA SAPIENZA.
XX
PA
XX Macino G, Cogoni C;
XX WPI; 2000-579171/54.
XX N-PSDB; AAE65171.
XX Novel polynucleotide encoding a polypeptide which has a silencing activity and comprising a RNA-dependent RNA polymerase domain.
XX Claim 21; Fig 4; 48pp; English.
XX The present sequence is the Neurospora crassa QDE-1 protein. This protein has gene silencing activity. The qde-1 gene was isolated by mutational analysis of an al-1 transgenic strain. This strain had an albino phenotype resulting from post-transcriptional silencing of the endogenous al-1 gene. Reversion of this phenotype indicated a mutation in a silencing gene. The silencing gene, qde-1, could then be isolated. CC Modulation of qde-1 expression may be used to inactivate genes and to silence suppression of genes
XX
SQ Sequence 1402 AA;

Query Match 100.0%; Score 7397; DB 3; Length 1402;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNPITPRKNSVBEIIINRLNDYNLGLQCVDATTLTPHRRKELAESDEDFGRHDKIYRA 60
DB 1 NNPITPRKNSVBEIIINRLNDYNLGLQCVDATTLTPHRRKELAESDEDFGRHDKIYRA 60
QY 61 NFLFYWRKDDSLNQAEANFFIEAKAASNWVPKHAHDPDTLPWSKEPPRAATAGQWALQ 120
DB 61 NFLFYWRKDDSLNQAEANFFIEAKAASNWVPKHAHDPDTLPWSKEPPRAATAGQWALQ 120
QY 121 TVLLEVLNRFMPPPNNTTGRITLSPGSLSRPTSTNTKRDKEPANVTADPPKSLT 180
DB 121 TVLLEVLNRFMPPPNNTTGRITLSPGSLSRPTSTNTKRDKEPANVTADPPKSLT 180
QY 181 RSATGPPHGAALPLKFPDPVNTGKRSLESENINQCTKRAKGLSDNAAAAAPPVPI 240
DB 181 RSATGPPHGAALPLKFPDPVNTGKRSLESENINQCTKRAKGLSDNAAAAAPPVPI 240
QY 241 ASALDKVPTERRHANTRDPTATGHRADQVDSPTSGTSGSSVFSACRHNQSTQSSFE 300
DB 241 ASALDKVPTERRHANTRDPTATGHRADQVDSPTSGTSGSSVFSACRHNQSTQSSFE 300
QY 301 APPSQPREKRPVDATVFAGHLIESPSKGRITTKSHIDNQPLSSSQGSETSTYVESPPS 360
DB 301 APPSQPREKRPVDATVFAGHLIESPSKGRITTKSHIDNQPLSSSQGSETSTYVESPPS 360
QY 361 SGGECAIPEPSRNSGLARSEARSQVQVHAPVVAARLNTWPKPKWLHEAPLAVAVEV 420
DB 361 SGGECAIPEPSRNSGLARSEARSQVQVHAPVVAARLNTWPKPKWLHEAPLAVAVEV 420
QY 421 TRLFVHCKVDLEDSIGLYKVDPSWSTARDVTDIWKTLVLDAPRGKPEKPPNDVFTA 480
DB 421 TRLFVHCKVDLEDSIGLYKVDPSWSTARDVTDIWKTLVLDAPRGKPEKPPNDVFTA 480
QY 481 MTGNFESKGSVAVLSAVLDYNDPNSPTAPLYLVKLKPLMFEQGCLTRRFGDFRFFELI 540
DB 481 MTGNFESKGSVAVLSAVLDYNDPNSPTAPLYLVKLKPLMFEQGCLTRRFGDFRFFELI 540
QY 541 PPTSTSPVPPVSKQCAVEVITQWLTMGHSLVGRQWRAFFAKDAGYKPLREFOLR 600
DB 541 PPTSTSPVPPVSKQCAVEVITQWLTMGHSLVGRQWRAFFAKDAGYKPLREFOLR 600
QY 601 AEDKPEIIEKRVHFFAETGITRPPDVFKTRSVPVPAEFPVEQTEPKVQMLDLWLLQDNN 660
DB 601 AEDKPEIIEKRVHFFAETGITRPPDVFKTRSVPVPAEFPVEQTEPKVQMLDLWLLQDNN 660
QY 661 TWQPHLKLSRIQLGLSKTYA:MTLEPHQIRHHKTDLLSPSGTGEVMDGVGRXSRSAK 720
DB 661 TWQPHLKLSRIQLGLSKTYA:MTLEPHQIRHHKTDLLSPSGTGEVMDGVGRXSRSAK 720
QY 721 RIRDVLGLGDVPSAVQGRFGSAKGMWIDVDTGDEDMTETVPSQKWECDVYDKHQRTL 780
DB 721 RIRDVLGLGDVPSAVQGRFGSAKGMWIDVDTGDEDMTETVPSQKWECDVYDKHQRTL 780
QY 781 EVRSVASELKSAGLNQLLPVLEDRARDKVRQRAIGDRLINDLQRFSEQKHANRPVE 840
DB 781 EVRSVASELKSAGLNQLLPVLEDRARDKVRQRAIGDRLINDLQRFSEQKHANRPVE 840
QY 841 FROWVYESYSSRATRVSHGRVFPFVPLAGLPSQSETNFMNSGDFPKKQYLDIANDLQK 900
DB 841 FROWVYESYSSRATRVSHGRVFPFVPLAGLPSQSETNFMNSGDFPKKQYLDIANDLQK 900
QY 901 RKCDTLKSLNIRVGRSAIYMIADFWGLVEENVEHVGFSSKFRDEESFTLLSDCDVLY 960
DB 901 RKCDTLKSLNIRVGRSAIYMIADFWGLVEENVEHVGFSSKFRDEESFTLLSDCDVLY 960
QY 961 ARSPAHPSPDIQVRAVAFKPELHSLKDVIIIFSTKGDVPLAKLSGDDYDGMWVCWDPE 1020
DB 961 ARSPAHPSPDIQVRAVAFKPELHSLKDVIIIFSTKGDVPLAKLSGDDYDGMWVCWDPE 1020
QY 1021 IVDGFVNAEMPLEPDLISRLYKDKKTTFKQLMASHGTGSAKSTQTYDMIQKSFHPALQFN 1080
DB 1021 IVDGFVNAEMPLEPDLISRLYKDKKTTFKQLMASHGTGSAKSTQTYDMIQKSFHPALQFN 1080

QY 1081 FLGMCNTYKERLCYINNVSNNKPAIILSSLVGNLVDQSKQGVFNESWAQLRRELGG 1140
DB 1081 FLGMCNTYKERLCYINNVSNNKPAIILSSLVGNLVDQSKQGVFNESWAQLRRELGG 1140
QY 1141 LSLPDPMYKSDWMLGRGEPHTIIDYLFKFSIARPAIDKELEAFHNAKAAKOTEDGAHFW 1200
DB 1141 LSLPDPMYKSDWMLGRGEPHTIIDYLFKFSIARPAIDKELEAFHNAKAAKOTEDGAHFW 1200
QY 1201 PDLASVYTFPKELISDKSRSSALLFTLKNRIGEVKEYSGRLVKNKEMRDSKDPYPVRV 1260
DB 1201 PDLASVYTFPKELISDKSRSSALLFTLKNRIGEVKEYSGRLVKNKEMRDSKDPYPVRV 1260
QY 1261 VYEKWCATITPEAMDKSGANYDSKVIRLLELSFLADREMTWALLRASTAFKLYYHKSP 1320
DB 1261 VYEKWCATITPEAMDKSGANYDSKVIRLLELSFLADREMTWALLRASTAFKLYYHKSP 1320
QY 1321 VQWMAQROLAYIAKQMTSRPGEAPALMTAFVYAGLMPDKKTKQYVARLEGDSGEYDP 1380
DB 1321 VQWMAQROLAYIAKQMTSRPGEAPALMTAFVYAGLMPDKKTKQYVARLEGDSGEYDP 1380
QY 1381 EYVEVLGDDDDFDGIGTGNGDY 1402
DB 1381 EYVEVLGDDDDFDGIGTGNGDY 1402

RESULT 2
AAG67235

ID AAG67235 standard; protein; 1196 AA.

XX AC AAG67235;

XX DT 13-NOV-2001 (first entry)

XX DE Amino acid sequence of a plant SGS2 polypeptide.

XX KW SGS2; RNA-dependent RNA polymerase; transgene silencing;
transgene stability; crop plant; viral resistance.

XX OS Arabidopsis thaliana.

XX FN FR2804128-A1.

XX PD 27-JUL-2001.

XX PF 26-JAN-2000; 2000FR-00001007.

XX PR 26-JAN-2000; 2000FR-00001007.

XX PA (RHOB-) RHOBIO.

XX PA (INRG) INST NAT RECH AGRONOMIQUE.

XX PI Beclin C, Elmayer T, Mourrain P, Vaucheret H;

XX DR WPI; 2001-543303/61.

XX DR N-PSDB; AAH77704.

XX PT A new plant SGS2 gene involved in encoding an RNA-dependent RNA
polymerase and in transgene silencing, increases transgene stability and
expression in transgenic plants when it is inactivated.

XX PS Claim 24; Fig 1; 46pp; French.

XX CC The present sequence represents a plant SGS2 polypeptide. SGS2 is an RNA-
dependent RNA polymerase and is involved in transgene silencing.
XX CC Inactivation of SGS2 is used to increase transgene stability and
expression in plants, particularly crop plants, especially maize, corn,
XX CC barley, sorghum, soy, sugar cane, beet, tobacco or cotton plants.
XX CC Overexpression of SGS2 can be used to increase resistance to viral
infection in plants

XX SQ Sequence 1196 AA;

Query Match

4.4%; Score 328; DB 4; Length 1196;

Best Local Similarity 22.7%; Pred. No. 3 6e-17;
Matches 197; Conservative 126; Mismatches 312; Indels 234; Gaps 42;
QY 535 FFEILIPSTSTSPSPVPVQPGAVEVIOIWMQSHVGVROWRAFFAKDGYRKP 594
DB 466 FVAPIVKDLTSSFSQKTVFVRVSI-----LTDG-FKLCGRKY-SFLAFSAN----- 512
QY 595 REFQRAEDPKPIIKERVHFAETGTTREPDPVKFTRSVVPAEPEVQTEFKVSOQLMDWL 654
DB 513 ---QLR-----DESANFFAEDG-----KTR-----VSDIKTWM 537
QY 655 LQL-DNNTWQPHKLFPSRIQLGLSTYAIMTLEPHQIRHHKTDLLSPSGTGEVMDGVGR 713
DB 538 GFKDKNV-----AKAARMGLCFSSYATVDVMPHEV---DTEVPDIERNGVFSGDIGT 590
QY 714 MSRSVAKIRIDVLGLGDV---PSAVQGRFGSAKGMWIVDVTGDGEDMIETVPSQ----- 765
DB 591 ITPDLADEVWEKLI-DVHYSFCAQIRYAGFGV-----VAWPSKSDGIR 636
QY 766 ---RKWECDFVDKHORTLEVRVSELKSAAGINLQLLPVLEDRARDKVMKQAIGDRLIN 822
DB 637 LALRDSMKKFFSKH-TILEICS-WTRFQPGFLNRQIITLLS-----VLGVDPDEIFW 685
QY 823 DLQRFSEQKHALNPVRQWVPSYSSRATRVSHGRVFPFLAGLPDSQEBTLNFMNSG 882
DB 686 DNQ-----ESMLYKLNRLDDTVAFEVLTASCA-----EQGNTAAILMSAG 727
QY 883 FPKKQKYLQDIAMDLQKCKDCLKNIRGRSAYITMIADFWGVLEENEVHVGFSS-- 940
DB 728 FPKTEPHLRGLMSSVRIAQLWGLREKRSIFVTSGRMLMGCLDEAGILEHGQCFIQVSKP 787
QY 941 -----SKFRDEESFTLLSCDVLVARSAPHFSDIORVAVFKPELHSLKDVII 990
DB 788 STENCFSKHSRFFKTKTDLVVKGY-VAIAKNPCLPDGVRAILEAVDPQLHMYDCLI 846
QY 991 FSTKGVPLAKKLSGGDYGDMWVCMDEIVDGFVNAEMPLEPDLRSY--LKDKTKTFK 1048
DB 847 FPKGDRPHITNEASGSLDGLDLYFVANDQKLI-----PPNKSYPAMHYDAEEK 896
QY 1049 QLMASHGTGSAKEQTYDMIOKSHFALQPNFLGCTN----YKERLCYINNSVSNKPA 1104
DB 897 SL-----GRAVNHODIIDFFARN-----LANEQLGTICNAHVHADRSY---GAMDEEC 943
QY 1105 IILSSIVGNLVDQSGKGVFNEASWAQLRRELLGGALSIPD--PMYKSDSLGR----- 1156
DB 944 LLLAELAAVDPFKTKIVSWP--PHLKPDPYPMGKEDVQTYKSNKILORLYRRVKE 1001
QY 1157 -----GEPHTI-----IDYLFSTARP-----AIDKELEAFHNAKAAKD 1191
DB 1002 VYDEDAEASSESTPSAIPYDAVLEIPGFEDLIPEAMGHKCLYDGLIGLLGQYKVKQE 1061
QY 1192 TE-DGAHFWDDPLASYTFKFKISDKSRSSALLFTTLKNRIGEVKEKRLVKNKEMRDS 1250
DB 1062 EBITVCHWS---MPKYTSKKQGEUKER-----LXHSYNSLKKEP-RKVFRETIPDH 1109
QY 1251 KDPYPRVNVQVYEK---WCAIT--PEAMDKSGANYDSKVIRLLELSFLADREMTWALL 1304
DB 1110 ENLSBEEKNILEKASAWHYTHYHPWYKGS-----LELQ---DPDESSHAAM 1155
QY 1305 RASTAFKLYHKSPEKVMQAGROLAYIK 1333
DB 1156 LS-----FAW-IAADYLARIK 1170

RESULT 3
AAB28532
ID AAB28532 standard; protein; 835 AA.
XX AC
XX AAB28532;
XX DT
XX 07-FEB-2001 (first entry)
XX DE Rice RNA-directed RNA polymerase, SEQ ID NO: 10.

XX Rice; plant: RNA-directed RNA polymerase; gene mapping; gene marker;
KW Plant virus resistance; plant breeding.
XX Oryza sativa.
XX WC2000060097-A1.
XX 12-OCT-2000.
XX 06-APR-2000; 2000WO-US009105.
XX 07-APR-1999; 99US-0128094P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
XX Zhong G;
XX WPI; 2000-679376/66.
XX N-PSDB; AAC63741.
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance.
XX Claim 10; Page 46-49; 62pp; English.
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX isolated from plant cDNA libraries. They are useful as probes for
XX genetically and physically mapping genes, and as markers for traits
XX linked to those genes. They are useful for controlling gene expression
XX and provide mechanisms to engineer plant virus resistance. They are also
XX useful for plant breeding to develop lines with desired phenotypes
XX Sequence 836 AA;
Query Match 4.3%; Score 318; DB 3; Length 836;
Best Local Similarity 21.1%; Pred. No. 1.3e-16;
Matches 225; Conservative 162; Mismatches 352; Indels 328; Gaps 52;
QY 376 LARSEARSQVQHPVVAARLNITW-PKFPKWLHEAPLAVAVEVTRLFMHC----- 427
DB 1 MXSEEN-----VNVALRDFWGDKFP-----VFACGRLLKAL 34
QY 428 -KYDLEDESGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPP-----N 474
DB 35 NRVARNPKLCSKVGDDHAEVRSVITPTXAYCL-----PPEVERSNRVLRHVH 83
QY 475 DVFTWMTGNFESGSAVLSAVLDYNPDNSPTAPLYLVKLKELMFEQGCRLTRRFGPDR 534
DB 84 EVADRFLRVTFMDEGMQVNNVNL-----NSFTAPI-----VKDLM-----SN 121
QY 535 FFEILIPSTSTSPSPVPVQPGAVEVIOIWMQSHVGVROWRAFFAKDGYRKP 594
DB 122 FFO-----QKTTVYKVRMLLTGPH-MCGRKY-SFLAFSSN----- 156
QY 595 REFQRAEDPKPIIKERVHFAETGTTREPDPVKFTRSVVPAEPEVQTEFKVSOQLMDWL 654
DB 157 ---QLR-----DKSANFFAE-----DKRT--TVEAIRKM- 180
QY 655 LQLDNNTWQPHKLF---SRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMDGV 711
DB 181 --MDGS---QVRMFEDAAARMGOCFSSTVATVMTRPDEVDSEFDDVVH---NEYIFSDGI 231
QY 712 GRMSRSVAKIRIDVLGLGD-VPSAVQGRFGSAKGMWIVDVTGDGEDMIETVPSQKWE 769
DB 232 GKITPDLAIEVAERLQTDNPPSAYQIRFAGFGK--VIAVWQHGSGDTLFLRPSMRKE 289
QY 770 CDFVDKHQRTLEVRVAS--ELKSAGINLQLLPVLEDRARDKVMKQAIGDRLINLQRO 827

Db 290 SNHL-----VLGVSWTKFQPGFLNRQIIILLSS-----LNVPDSIFWQMDET 332
QY 828 FSEQKHALNRPVEPQWVYESSYSSRATRVSHGRVPFLAGLPDSQEEITFLNMSGDFPKX 887
Db 333 MLN---LNNILSDRDVAFVLTSCA-----DONTAALMSAGFPRT 374
QY 888 QYLODIADWLQKRCDTLTKLNTRVGRSAYIYMIADFVGLBE-----NEV 935
Db 375 EPHLKAMLLAIRSAQLQDLLEKARIFVPEKRWLMGCLDELGVLEGGQCFIRATVPSLNSY 434
QY 936 HVGFSKFRDESEFTLLSDCDVLVARSPAHFPPSDIORVRAVPELHSLKDVLIIFSTKG 995
Db 435 FVKGSRSSSTDKNEVILG-TVIAKPCPLHGDVRIEADVPELHVLVDCVFPQKG 493
QY 996 DVPLAKLUGGDDYDGMWVCDPEIVDFVNAEMPLE---PDLSRYLKDKDTTFKQLMA 1052
Db 494 ERPHANEASGSLDGLYFVTWDEKLIPPGKKSNWPMYSPPEA-----KGL-- 540
QY 1053 SHGTSAAKEQTYDMIQSHFALQPNFLGCTNYKERLCYINNSVSNKPAITLSSLVG 1112
Db 541 -----PRQVSHDIIDFFLKNMISENLGRICNAHVHVLSEYGVAMDEKCIHLAELAA 593
QY 1113 NIVDOSKOGIVNEASWALRRELGGALS---LPDPM-----YKSDSWLGR----- 1156
Db 594 TAVDFPKTG-----KLAIWPHLPKPVYDFWKGEDGQSYKSEKILGLRYSIQ 642
QY 1157 ----CEPHIIDLYKFSIARP-----AIDKELEAFHNAKAKTEDGAHFWDPDLASY 1207
Db 643 EASNG-----DVVSQEVCTNDLYDIDLEVPASDFLASWQCKS---YDQLSALL 693
QY 1208 TPFKEISDKSRSSALLFTTLKNRICEVEKEYGRVLYKN--KEMRDSKDPYPVRVNVQVYEKW 1265
Db 694 SQY-----RVRTAEALVT-----GHIT-----FLVKNSSKKQGDIKO---RLKTA 731
QY 1266 CAITPEAMDKS-GANYDSKVRLLELSFLADREMTWALLRASTAFKLYHKSPKFWQM 1324
Db 732 ----SALRKEFKFSTFIASDQCE---IGDDEKNLLYEMKASAWYQVTH--PKWVEKS 781
QY 1325 AGRQLAYIKAQMTSPGCGAPALMTAFMYAGLMPDKKFTKQYVARLE 1371
Db 782 RG-----ILGPDGGEIFASLS---FAMIPVD-----YLARIK 810
RESULT 4
ID AAU10006 standard; protein; 1114 AA.
XX AAU10006;
XX AC AAU10006;
XX DT 08-MAY-2002 (first entry)
XX DE Tomato RNA-directed RNA polymerase (RdRP) protein.
XX KW Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW transgenic plant; transgenic animal; cancer; viral infection;
KW immunoprecipitation; immunolocalisation; EC 2.7.7.48; gene therapy.
XX OS Lycopodium esculentum.
XX PN US2001023067-A1.
XX PD 20-SEP-2001.
XX PF 08-FEB-2001; 2001US-00782874.
XX PR 05-MAR-1997; 97US-00811583.
XX PA (WASS/) WASENEGGER M.
XX PA (RIED/) RIEDEL L.
XX PI Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX

DR WPI; 2001-595798/67.
XX N-PSDB; AAS17837.
PT New nucleic acid molecule encoding a polypeptide having the enzymatic
PT activity of RNA-directed RNA polymerase, for modulating gene expression
PT and treating cancer and virus infection in human and animals.
XX PS
XX Claim 1; Page 19-22; 34pp; English.
XX This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
CC protein of the invention. The invention comprises the nucleic acid and
CC protein sequences of RdRP. The protein of the invention can catalyse in
CC vitro transcription of short single stranded RNAs into DNA molecules,
CC this transcription can be either primed by RNA or DNA oligonucleotides or
CC be unprimed. The protein may have cytostatic or virucide activities. The
CC sequences of the invention may be used in gene therapy or as an RNA
CC directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
CC nucleic acid molecule derived from a nucleic acid molecule which causes a
CC disease are useful for treating a disease caused by the undesired
CC expression or overexpression of a nucleic acid molecule in a human, rat
CC or mouse, by administering the molecules. This system can be used in the
CC preparation of a pharmaceutical composition and for inhibiting expression
CC of any desired gene by transferring the RdRP system to organisms that
CC either lack a comparable mechanism or do not sufficiently express their
CC own RdRP. An antibody or an antagonist or inhibitor to the protein are
CC useful for inhibiting RNA directed RNA synthesis and for ensuring stable
CC heterologous, gene expression in transgenic organisms. The sequence is
CC useful for probes and/or for the control of gene expression, as primers
CC for amplification of nucleic acid molecules and as tools for the
CC detection of expression of the cDNA molecules. Additionally, nucleotide
CC and protein sequences are useful for suppression of undesired gene
CC expression in humans and animals. The RdRP is useful as a therapeutic
CC agent for the control of cancer and virus infection in humans and animals
CC and the antibody is useful for immunoprecipitation or immunolocalisation
CC of the protein, identification of polypeptides interacting with it and
CC screening expression libraries
XX Sequence 1114 AA;
Query Match 4.0%; Score 297; DB 4; Length 1114;
Best Local Similarity 20.4%; Pred No. 1.2e-14;
Matches 175; Conservative 133; Mismatches 323; Indels 228; Gaps 32;
QY 562 BEVIOWLTGQHSVGRQWRAFFAKDAGYRKPLRFQRAEDPKPIIKERVFFETGTT 621
Db 386 EDIDNFLRV---SFVDEWEKLYSTDL-----LPK-----ASTGSG 418
QY 622 FRPDVEK-----TRSVPAEPEPQRTQEFKYSQMLDMLQLDNNTW---OPHLKL--- 668
Db 419 VRTNIYERILSTLRKGFVIGDKKFE-FLAFSSQL-----RDNVWMEASRPGTANDI 471
QY 669 -----FSRI-----QLGSKTYAIMTLEPHQIRHHKTDLL---SPSGTEVWMDGV 711
Db 472 RAMWGDPSQIKNVAKYAARLQSGFSSRETL---SVLRHEIEVIPDVKVHGTSVFSDGI 528
QY 712 GRMSRVAKEIRDVLGLGVPSAVQGRGSAKGMVVDVDDTGDDEWIEITYPSQRKWECD 771
Db 529 GKISGDFARVASKCGLQYTPSAFQIRYGGYGVGVDPDSSNK---ISLRKSMKSYSD 585
QY 772 FVDXHQRTLEVRVASELSAGLNLQLLPVLEDRADRVKMQEQAIGDRLNDLQRFSSQ 831
Db 586 NI-----KLDVLG-WSKYQPCVNLNQLITLS-----TLGVKDEVLQKQKEAVDQ 630
QY 832 KHALRNPVEFRQWVYESSYSSRATRVSHGRVPFLAGLPDSQEEITFLNMSGDFPKKQYL 891
Db 631 LDAILHDSLKQAQFAELMS-----PGENTNMLKAMLMCGKPAEPFL 673
QY 892 QDIAMDLOKRCDDTLTKLNIRVGRSAYIYMIADFVGLBE-----SKFRDEE 947
Db 674 SMMLOTFRASKLLDLRTSRIFIPNGRTWMGCLDSRTLEYQGVVQFTGAGHGFSDDL 733
QY 948 ESE-----TLLSDCDVLVARSPAHFPPSDIORVRAVPELHSLKDVLIIFSTKGDVPL 999

| | | | |
|----------|---|---|------|
| Db | 734 | HPFNRSRSTNSFILKGNVVAKNPCLHPGDIRVLKAVNRALHMHVDCVFPQKGRPH | 793 |
| Qy | 1000 | AKKLSGGYDGDMAVWCVDPVIDGVNAEMPLEPDLRYLKKDKTTFKQLMASHGTGSA | 1059 |
| Db | 794 | PNECSGSLDGDYFVVCWQDMIPPRQVQPMYPPAPSIQDHD | 837 |
| Qy | 1060 | AKEQTYDMIQSFHALQPNFLGMCTN---YKERLCYINNSVSNKPAIILSSLVGNLV | 1115 |
| Db | 838 | ---VTIEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAEIIFSIAV | 890 |
| Qy | 1116 | DOSKOGIVNEASWAQLRELLGGALSILPDPMYKSDSWLGRGEPHIDYIKFSIARPAI | 1175 |
| Db | 891 | DPFKTGVFAETPS---QLSPK-----EYDFMDKPD-----KTSYISERVIGLFFKVK | 936 |
| Qy | 1176 | DKELEAFNAKAAKDTEDGAHFWDPDLASYTTFFKEISDKSRSSALLFTTLKNRIGEV- | 1234 |
| Db | 937 | DKA-----PQASSIATFTRDVARSDYADMEVDGFEYIDEAF | 974 |
| Qy | 1235 | --EKYGLVKNKENDSKDYPVRVQVQYKWKCAITPEAMDKSGANYD-----SKV | 1284 |
| Db | 975 | DYKTEY-----DNKLGNDMYGKTE-----AEI-SGGIMKASKTFDRRKDAEALSVA | 1023 |
| Qy | 1285 | IRLLELSFLADREMTW-----ALLRASTAFKLYYHKS-----PKFV | 1321 |
| Db | 1024 | VRALR-----KEARAWFKRNDIDDMPLKASAWHYTHPTYWGCVYNOGLKRAHFISFP | 1077 |
| Qy | 1322 | WOMAGROLAYTKAQWTSRP | 1340 |
| Db | 1078 | WCYDQLIQIKDKARNRP | 1096 |
| RESULT 5 | | | |
| ID | AAE00897 | standard; protein; 1114 AA. | |
| XX | AAE00897; | | |
| AC | AAE00897; | | |
| XX | AAE00897; | | |
| DT | 04-JUL-2001 | (first entry) | |
| XX | Tomato C-protein having RNA-directed RNA polymerase (RdRp) activity. | | |
| DE | Tomato; gene therapy; RNA-directed RNA polymerase; RdRp; gene expression; | | |
| XX | transgenic plant; tissue culture; plant breeding; therapy; C-protein. | | |
| KW | Lycopersicon esculentum. | | |
| XX | Location/Qualifiers | | |
| PH | 331..346 | | |
| FT | /label= P432 | | |
| FT | /note= "C-protein-specific antibody for immunisation of | | |
| FT | rabbits" | | |
| FT | 377..390 | | |
| FT | /label= P430 | | |
| FT | /note= "C-protein-specific antibody for immunisation of | | |
| FT | rabbits" | | |
| FT | 784..797 | | |
| FT | /label= P433 | | |
| FT | /note= "C-protein-specific antibody for immunisation of | | |
| FT | rabbits" | | |
| FT | 1007..1020 | | |
| FT | /label= P431 | | |
| FT | /note= "C-protein-specific antibody for immunisation of | | |
| FT | rabbits" | | |
| XX | US6218142-B1. | | |
| XX | 17-APR-2001. | | |
| XX | 05-MAR-1997; | 97US-00811583. | |
| PF | 05-MAR-1997; | 97US-00811583. | |
| XX | 05-MAR-1997; | 97US-00811583. | |
| PR | (WASS//) WASSENEGGER M. | | |
| XX | PA | | |

| | | |
|----|---|--|
| PA | (RIED//) RIEDEL L. | |
| XX | Wassenegger M, Riedel L, Schiebel W, Sanger HL; | |
| PI | WPI; 2001-289830/30. | |
| XX | N-PSDB; AAD04370. | |
| DR | New nucleic acid molecules encoding polypeptides with RNA-directed RNA | |
| PT | polymerase enzymatic activity, useful in modulating gene expression in | |
| PT | plants, humans and animals, as well as in plant cell/tissue cultures or | |
| PT | plant breeding. | |
| XX | Claim 1; Col 33-40; 31pp; English. | |
| PS | The present sequence is a tomato C-protein having RNA-directed RNA | |
| CC | polymerase (RdRp) activity. This protein is capable of RNA-directed RNA | |
| CC | synthesis, thus using RNA as a template for synthesising complementary | |
| CC | RNA molecules. RdRp nucleic acid is useful for modulating gene expression | |
| CC | in plants, humans and animals. This may lead to various physiological, | |
| CC | developmental and/or morphological changes. Transgenic plants containing | |
| CC | RdRp nucleic acid is especially useful in plant cell or tissue cultures | |
| CC | and in plant breeding. RdRp is useful in gene therapy, particularly for | |
| CC | treating a disease that is caused by the undesirable expression or | |
| CC | overexpression of a gene | |
| XX | Sequence 1114 AA; | |
| Qy | Query Match 4.0%; Score 297; DB 4; Length 1114; | |
| Db | Best Local Similarity 20.4%; Pred. No. 1.2e-14; | |
| Qy | Matches 175; Conservative 133; Mismatches 323; Indels 228; Gaps 32; | |
| Db | 562 EVIQWLTMGHSLVGRQWRAFFAXDAGYRKRLRFQRAEDPKLIKRVHFFETGIT 621 | |
| Db | 386 EDIDNPLRV---SFVDEWEKLYSTDL-----LPK-----ASTGSG 418 | |
| Qy | 622 FRPDVPK-----TRSVVPABEPVQRTFVKYSQMLDWLQLDNNTW---QPHLKL--- 668 | |
| Db | 419 VRTNIVERLTSLRKGFVIGDKKFS-FLAFSSQL-----RDNSVWMPASRPLGTANDI 471 | |
| Qy | 669 -----PSRI-----QLGSKTYAINTLPHQIRHKTDL---SPSGTGEVWANDGV 711 | |
| Db | 472 RAMGDFSGIKNKVAKYAARLGQSGSSRETL---SVLRHEIEVI PDVKHGSYVFSOGI 528 | |
| Qy | 712 GMSRSVAKIRIDLGLGDPVSAVQGRFGSAGMWDVDDTGDDETWETETYSQKWECD 771 | |
| Db | 529 GKISGDFAHVAVASKGLQYTPSAFQIRYGGYKGVGVDPDSSMK---LSLRKSMKYESD 585 | |
| Qy | 772 FVDKQRTLEVRVASELSKAGLINLQLLPVLEDRARDKVRQRAIGDRINDLQRFSEQ 831 | |
| Db | 586 NI-----KLDVLG-WSKYQPCVLRQLITLLS-----TLGVKDEYLEQKQEAVDQ 630 | |
| Qy | 832 KHALARPVEFRQWVYESYSSRATRVSHGRVPLAGLPDSQEBTLNFMNSGDFPKKQYL 891 | |
| Db | 631 LDAILHDSLKAEALMNS-----PGENTNLIKAMLCNGKYPDAEPFL 673 | |
| Qy | 892 QDIAMDQKRKCDTLKSKLINRVGRSAIYMIADFWGLNEVHVGFS---SKFRDEE 947 | |
| Db | 674 SMMLQTFRASKLLDRTSRIFIPNGRTMWGCLDESRTLEYQVQVQFTGAGHGFSDDL 733 | |
| Qy | 948-ESF-----TLLSDCQVLVARSAPHFSPDIQVRVAFKELHSLKVLIFSTGVDVPL 999 | |
| Db | 734 HFNNSRSTNSNFIKGNVVAKNPCLHPGDIRVLKAVNRALHMHVDCVFPQKGRPH 793 | |
| Qy | 1000 AKKLSGGYDGDMAVWCVDPVIDGVNAEMPLEPDLRYLKKDKTTFKQLMASHGTGSA 1059 | |
| Db | 794 PNECSGSLDGDYFVVCWQDMIPPRQVQPMYPPAPSIQDHD----- 837 | |
| Qy | 1060 AKEQTYDMIQSFHALQPNFLGMCTN---YKERLCYINNSVSNKPAIILSSLVGNLV 1115 | |
| Db | 838 ---VTIEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAEIIFSIAV 890 | |
| Qy | 1116 DOSKOGIVNEASWAQLRELLGGALSILPDPMYKSDSWLGRGEPHIDYIKFSIARPAI 1175 | |

Db 891 DFPKTVPAEIPS--QLRPK-----EYDFNDKPD-----KTSYISERVIGLFRKVK 936

Qy 1176 DKELEAFNMAKAAKDTEDGAHFDPDLASYITFFKEISDKSRSSALLFTTKNRIGEV- 1234

Db 937 DKA-----PQASSIATFTRDVARRSYDADMEVDGFDYIDEAF 974

Qy 1235 --EKEYGRLVKKEWRDSDKDPYVRNVQVYKWKCAITPEAMDKSGANYD-----SKV 1284

Db 975 DYKTEY-----DNKLGNLMDYGIKTE-----AEILSGGIMKASKTFDRKDAEASVA 1023

Qy 1285 IRLLELSFLADREMTW-----ALLRASTAFKLYYHKS-----PKFV 1321

Db 1024 VVALR-----KEARAFKERNDDIMLPKASAWHYVHYHPIYWCYNQGLKRAHFISFP 1077

Qy 1322 WQVAGROLAYIKAQMTSRP 1340

Db 1078 WCVYDQLIQIKKAKARNRP 1096

RESULT 6

AAB28531

ID AAB28531 standard; protein; 1164 AA.

XX AAB28531;

AC

DT 07-FEB-2001 (first entry)

XX

DE Maize RNA-directed RNA polymerase, SEQ ID NO: 8.

XX

KW Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;

KW plant virus resistance; plant breeding.

XX

OS Zea mays.

XX

XX WQ200060097-A1.

PN

XX 12-OCT-2000.

PD

XX

XX 06-APR-2000; 2000WO-US009105.

PF

XX

XX 07-APR-1999; 99US-0128094P.

PR

XX

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

PA (FION-) PIONEER HI-BRED INT INC.

PA

XX

PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;

PI Zhong G;

PI

XX

XX WPI; 2000-679376/66.

XX

DR N-PSDB; AAC63740.

DR

XX

XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful

PT for controlling gene expression and providing mechanisms to engineer

PT plant virus resistance.

XX

XX Claim 10; Page 41-45; 62pp; English.

PS

XX

CC The present sequence is part of a plant RNA-directed RNA polymerase

CC protein. Polynucleotides encoding RNA-directed RNA polymerases were

CC isolated from plant cDNA libraries. They are useful as probes for

CC genetically and physically mapping genes, and as markers for traits

CC linked to those genes. They are useful for controlling gene expression

CC and provide mechanisms to engineer plant virus resistance. They are also

CC useful for plant breeding to develop lines with desired phenotypes

XX

XX Sequence 1164 AA;

SQ

Query Match 3.9%; Score 292; DB 3; Length 1164;

Best Local Similarity 21.4%; Pred. No. 3.5e-14;

Matches 176; Conservative 123; Mismatches 323; Indels 200; Gaps 35;

Qy 605 KPIIKERVHFAETG-----ITFRPDVKTASVVPABEFVQRTFEKYSQMLDWL 654

Db 453 KTTVYKRVKTFLTGEGHMGCRKYSFLAFSSNQIRDRS---AWFFAEDRTT-TVETIRKWM 508

Qy 655 LQDNNNTWPHLKLFSRIQLGHSKTYAINTLPHQIRHHKTDLLSPSGTGEVWNDGVGM 714

Db 509 GRFTSKNVAKHA---ARMGQCFSSYATVVLQPHVEVNECLDEV---EHNGYIFSDGIGKI 562

Qy 715 SRSAVAKIIRDVLGLG-VPASVQCRGFSAGKMWIVDV-DTGDDEWIEYTPSKWECDP 772

Db 563 TCDLALVLAQKLOLTNPSSAIVIRYAGFKG-VTSWEGKNDGIRLSRPSMHKE--- 617

Qy 773 VDXHORTLEVRSAVSKSAGLNQLLPVLEDRARDKVRQRAIGDRRLINDIQRQFSEOK 832

Db 618 --SNHTVLEVS-WTKFQGFNLNRQIITLLS----- 645

Qy 833 HALNRVPEFRQWVYESYSSRATFV-SHGRVFF--LAGLPDSQEBETLNLMSGDFDKKOK 889

Db 646 -SLNVDAIPAQMEAWLNLNLSDSVADIVTASCACQTTAALMLSAGISFSTEP 704

Qy 890 YLQDIAMDLOKRCCKDLTKSKLINRVSAYIYMIADFWGLBENEVHVHVSFKPRDEES 949

Db 705 HLKAMLLAIRSSQLLLEKTRIFVFKRWLMGCLDELGLEQGCQFIRASS----- 756

Qy 950 FTLLSDC-----DVLVARSAPHFPSDIOQRVAVFKPELHSLKDV 989

Db 757 -PSLNCLVKYGRSLSAANTNAETILGTIVMAKNPCLPCHFGDVRILEAVDPVPELHLVDCL 815

Qy 990 IFSTKGDVPLAKKLGGDYDGMWAVYCDPEIYDGFVNAEMPLEPDLRYLKKDKTTFKQ 1049

Db 816 VPEKKGSRPHANEASGSLDGLDYFTVDENLIPPGCKSWNPMD-----YSPAFAKQLPR 870

Qy 1050 LMASHGTGSAKEQTYTDMIQKSFHALQPNFLGMCNTYKERLCYINNS----- 1098

Db 871 AVSQH-----DIV--GF-----FLKNNVN--EKLGPISNHVVHADNSEVG 907

Qy 1099 VSNKPAILSSLVGNLVDSKQIGIVFNEASWAOQLRELLGGALSPLDPM-YKSDSWLGR- 1156

Db 908 AMDEKCIQLAELAATAVDFFKTGIYSMP--ASLRPKLYPDFMGKEDAISYRSEKILGRL 965

Qy 1157 -----GEPTHIIDLYKFS-IARPAIDKELEAFNAMK 1187

Db 966 YRSIQEASDDLVPETCTSNLFPYDADMEVAGADFLSAMOCKSVETQLNALLNQYG 1025

Qy 1188 AAKDTE-DGAHFWDPDILASYITFFKEISDKSRSSALLFTTKNRIGEVEKEYGRLVKNE 1246

Db 1026 VRTEAEIVTEHIW--SLPKY-----SSRQGDIKERLKNAYVALHKEF-RSIFESI 1073

Qy 1247 MRDSKDPYVRVNVQVYE---KWCAIT--PEAMDKSGANYDSKVIRL-LELSFLADREYN 1299

Db 1074 VTDQTEISDDEKSRFVEMKASAWYQVYTHPEWYQKREMPKSCDEMPARLSF----- 1126

Qy 1300 TWALLRASTAFKLYYH-----KSPKFWQVAGRLAYIKAQM 1336

Db 1127 ANIAVEHLARIKIRCGEVKVDSPRPVERLA---AYISGSM 1164

RESULT 7

AAB28533

ID AAB28533 standard; protein; 1125 AA.

XX

AC AAB28533;

XX

DT 07-FEB-2001 (first entry)

XX

DE Soybean RNA-directed RNA polymerase, SEQ ID NO: 12.

XX

KW Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;

KW plant virus resistance; plant breeding.

XX

OS Glycine max.

XX

PN WQ200060097-A1.

XX

PD 12-OCT-2000.


```

XX PS Claim 10; Page 36-40; 62pp; English.
XX
CC The present sequence is part of a plant RNA-directed RNA polymerase
CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC isolated from plant cDNA libraries. They are useful as probes for
CC genetically and physically mapping genes, and as markers for traits
CC linked to those genes. They are useful for controlling gene expression
CC and provide mechanisms to engineer plant virus resistance. They are also
CC useful for plant breeding to develop lines with desired phenotypes
XX
SQ Sequence 1108 AA;
Query Match 3.6%; Score 267.5; DB 3; Length 1108;
Best Local Similarity 21.7%; Pred. No. 3.5e-12;
Matches 155; Conservative 113; Mismatches 304; Indels 141; Gaps 26;
QY 645 FKVSQMLD--W-----LLQDNNWQPHUKLFSRIQ-----LGLSKTYAINTLEPHQ 689
DB 465 FSSSQLRDNASAWFASRQGLTASDIRKWMGDFRDRNVAKYAAAILGQSFSSSTETLVHK 524
QY 690 IRHKTDLLSPSGTGEVNDGVRMSRSVAKRIQVGLGD-VPSAVQGRGSAKGMWVI 748
DB 525 SEVERIPDIT-NGTKYIFSDGGVKISANFAVEAMKCKLPAPSVQIRGGYGVAV 583
QY 749 DVDDTGDDEWIETPSQRKWECDVYDKHQTLEVRSAVELSKAGLNLQLPVLDRARD 808
DB 584 ---DTRSNHKLRLKSMKFSQENI-----TLDVLAY-SKYQPCFLNQLITLIS----- 629
QY 809 KYMKRQAIGRLINDLQRFSEQKALNRPVEFRQWVYESSRATRVSHGVRPFLAGLP 868
DB 630 ----TLGVSNDVPELKQKALRQ---LNRMT-----EPQAREAVELMPM----- 668
QY 869 DSOEETLNLMSGFPKQKYLQDIAMLQKRDITLKSINIRVGRSAIYMIADPWG 928
DB 669 GEVTNVKLLSGYQPDHEPVLMLQTFRASKLELTKERIITQGRAMMGLDFTC 728
QY 929 VLENEVHVGFSGKFRDEESTLLSDCDVLVARSAPHPSDIQRVAVRFPKLHSLKDV 988
DB 729 TLKYGQVFQASYSADDRHK---VVVTGKVVVAKNCPCLHPGDIRVLQAVDPALHFLFDC 785
QY 989 IIFSTGQVPLAKLGGGVDGDMWVCDPDIQVDFVNAEMPLEPDLRSVLKDKDTTFK 1048
DB 786 VVFPQOGPRHPNECSGLDGDIFVSWDPLHPSRLVDPMDYTPASAEITLDDH----- 840
QY 1049 QLMASHGTGSAKEQTYTDMIOKSFHALQPNFLGMCVNY-----KERLCYINNSVNK 1102
DB 841 -----VTBIEIQEYFTNYVNESLGIANAHVVFDDQERM-----KAESE 880
QY 1103 PAIILSSVLGNLVDQSKOGIVFNEASWAOELRLLGALSLEDDPM-----YKSDSWLG 1155
DB 881 PCVQLAKLPSIAVDFFKTGVP-----ALIPHLEL-HVKEYDFWEKLDKYTESKGVIG 932
QY 1156 R-----GEPHIIDYLFKSIARPAIDKEL--BAFNMAKAAKDTEDGAHFWDPDILASYTT 1208
DB 933 KLYREIKKHTPHIKHFTREARRSYDTDLIVDGYEDYITEAIEFFEYDFRLGNLMDHYG 992
QY 1209 FFKE---ISDKSRSSALLFT-----TLKNRICEVEKE---YGRLVKNKMRDSDKP-- 1253
DB 993 IKSEAIHISGCIUKMANFTKSSDADAIRNARSURKEARSFNENSTGEGQDAWEAKA 1052
QY 1254 ---YPRVNVQVY-----EKWCAITPEAMDKSGANDYSKVRILL 1288
DB 1053 SAWHYVTHQYNGSYNEGDRPHLISFFVCVYDKLVAKQGRNLLTQMDRNL 1105

```

RESULT 9

AAB28529

ID AAB28529 standard; protein; 417 AA.

XX

AC AAB28529;

XX

DT 07-FEB-2001 (first entry)

```

XX DE Maize partial RNA-directed RNA polymerase, SEQ ID NO: 4.
XX
XX KW Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
XX KM plant virus resistance; plant breeding.
XX
XX OS Zea mays.
XX
XX PN WO200060097-A1.
XX
XX PD 12-OCT-2000.
XX
XX PF 06-APR-2000; 2000WO-US009105.
XX
XX PR 07-APR-1999; 99US-0128094P.
XX
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX PA (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Odell JT, Orozco EM, Wang Z, Sakai H, Cahoon RE, Wang J;
XX PI Zhong G;
XX
XX WI: 2000-679376/66.
XX
XX DR N-PSDB; AAC63738.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX PT for controlling gene expression and providing mechanisms to engineer
XX PT plant virus resistance.
XX
XX PS Claim 10; Page 34-35; 62pp; English.
XX
XX The present sequence is part of a plant RNA-directed RNA polymerase
XX protein. Polynucleotides encoding RNA-directed RNA polymerases were
XX CC isolated from plant cDNA libraries. They are useful as probes for
XX CC genetically and physically mapping genes, and as markers for traits
XX CC linked to those genes. They are useful for controlling gene expression
XX CC and provide mechanisms to engineer plant virus resistance. They are also
XX CC useful for plant breeding to develop lines with desired phenotypes
XX
XX SQ Sequence 417 AA;
Query Match 3.2%; Score 236; DB 3; Length 417;
Best Local Similarity 21.3%; Pred. No. 2.7e-10;
Matches 84; Conservative 74; Mismatches 158; Indels 78; Gaps 11;
QY 598 LSPSGTGEVNDGVRMSRSVAKRIQVGLGDV--PSAVQGRGSAKGMWVIDDVTGD 755
DB 26 VTTDGTGYIFSDGIGKISTRFARQVAKLIGLDDPAHPSPAFQIRYGYKG--VITIDPVSF 83
QY 756 EDWETIYPSQRKWECDVYDKHQTLEVRSAVELSKAGLNL-----QLLPVLEDRARDKVK 811
DB 84 FN-LSLRPSMKPFES-----KSTMLNITNWSKSPQCVVNRHISLL 123
QY 812 MRQAIGRLINDLQRFSEQKHALNRPVFRQWVYESSRATRVSHGVRPFLAGLPDSQ 871
DB 124 STLGIKDEVFASQDDMHESDGLTNKE-----AALSVLGKIGGGD 165
QY 872 EETLNFNLMSGFPKQKYLQDIAMLQKRDITLKSINIRVGRSAIYMIADFWGVL 931
DB 166 TKTAADMLLGQYFSPSEPEYLLMLKARANRLDITRCKIHVKQGRVLIGCLDETCKLE 225
QY 932 ENEVHVGFSGKFRDE---EESF-----TLLSDCDVLVARSAPHPSDIQRVAVRFPKE 981
DB 226 YGQVYIRITKNRKEQKYSEQFFCNDDGKTAVTGVKVAITNFCPLHPGDIRVLEAVYDPG 285
QY 982 L--HSLKDVIIPTKGDVPLAKLGGDYGDMWVCDPDIQVDFVNAEMPLEPDLRSY 1039
DB 286 LDARGLIDCVVFPQGRHPRHPCSGGDLDDDDFFITWDDKLFPEKVDAPM----- 336
QY 1040 LKKDKTTFFKQLMASHGTGSAKEQTYTDMIOKSF 1073
DB 337 ---DYTATRPRIMDHA-----VTLEEIQKH 359

```


Chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.
Homo sapiens.
WO200153312-A1.
26-JUL-2001.
26-DEC-2000; 2000WO-US034263.
23-DEC-1999; 99US-00471275.
21-JAN-2000; 2000US-00488725.
25-APR-2000; 2000US-00552317.
20-JUN-2000; 2000US-00598042.
19-JUL-2000; 2000US-00620312.
03-AUG-2000; 2000US-00653450.
14-SEP-2000; 2000US-00662191.
19-OCT-2000; 2000US-00693036.
29-NOV-2000; 2000US-00727344.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren Z, Wang D,
Zhou P, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
Zhou P, Goodrich R, Drmanac RT;
WPI; 2001-442253/47.
N-PSDB; AA157835.
Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
Example 3; SEQ ID NO 1824; 10078pp; English.
The invention relates to human nucleic acids (AA157798-AA161369) and the
encoded polypeptides (AA138442-AA142213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as; immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders. Note: The sequence data for this patent did not form
part of the printed specification
SQ Sequence 966 AA;
Query Match 2.4%; Score 179.5; DB 4; Length 966;
Best Local Similarity 19.3%; Pred. No. 6.2e-05;
Matches 175; Conservative 133; Mismatches 333; Indels 265; Gaps 40;
QY 132 PPNNTPTGRTGLSGSLRPTSTNTKRKDEPA---NVTADPPKR-SLTSATGP 186
DB 65 PPP-----PEVPEVSTQPPPEPCGSETVPDGRSDSVSGSPRP 110
QY 187 PIHGAALPKFPDPVNTGSRPSLESENLN-----QCTKRAKGLSDNVAAA 234
DB 111 PSHFSFAVDEDKPI-ASSGTYNLDFDNIELVDTFQTLPRASDAKNOEGKV--NTRRKS 167
QY 235 APVPIA-SALDKVPTERRANTDPTATGHRADQVSDTSGTSGSVFSACR--- 289
DB 168 TDSVPISKSTLSLSLQASDFGASSGNPFAVALAPDAYGTGSSASSTLKRTKPRP 227
QY 290 ---HNSQTGSSPEAFPSOPREKRPVDATVFAAGHIESFKGRTTKSH-----IDNQP 340
DB 228 PSLKKQTTKPTFTFVKETQEPDEESLVPSGENLASETNTESAKTGSPALLEETP 287

QY 341 LSSSQGETSFSTYYESFPSSGGEGAIPEPSSNGLARSEESARSOVQVHAPVVAARLN 400
DB 288 LE-----FAVGPKAACLDSSAEGVPPASGGGRVQNSPPVGRKTL-- 329
QY 401 IMPFKPKLHEAPLAVAVEVTRLFMHCKVDLEDE-----SLGLKYD-----PSWSTARD- 449
DB 330 -----PLTTAPEAGEVTPSDSGQSDSPAKGLSVRLFDYFSDKSSWNDQOEN 377
QY 450 -----VTDIMKTLRLDAPRGKFFPEKP--PNDVFVTAMTGNFESKG 489
DB 378 PPPTKIGKPKVAMPLRRPKMKKTPKLDNTPASP-PRSPAEPNDIPAKGTYTID--- 433
QY 490 SAVVLSAVLDY--NPDNSPTAPLYLVKLKPLMFEGQCRLTRFGPDRFPEILIPSTST- 546
DB 434 -----IDKDDPNFNPFSSTSKQESFKLPQ-----SYNFDPTCDSDVPFKTSK 481
QY 547 SPSVPPVWSKQCAVEZEVQWLTWQHSLVG-----RQWRAFFAKDAGYKRP 593
DB 482 TPSSP---SKSPASFE--IPASANEANGVDGGLNKPAPKAKKTPKTDIFRVKKSFKRSP 536
QY 594 LREFQLRAEDP-----KPIIKERVHFFAETGI----- 620
DB 537 LSD--PPSQDPTPAATPETPPVISAVVHATDEKLAVTNQKWTMTVDLEADKQDYPQPS 594
QY 621 ---TFPPDVFKRSVVPAABEPVEQTEPKVQSM-----LDWLLQLDNNWTQ 663
DB 595 DLSTP---VNETKFSPTBE--LDYRNSYEIEYMEKIGSSLPODDAPKQKQALYLMFDTSQ 650
QY 664 PHLKPSRIQGLS-----KTYAIWT-----LEPHQIRHHKHTDLLSPS 701
DB 651 ESPVKSPPVMSPTSCSSSFEETALVNTAAKNQHPVPRGLAPNQESH---LQVPE 706
QY 702 GTGEVNDGVGRMSRVAKRIRDLVGLGV---PSAQGRFGSAK-----GMVV--IDVD 751
DB 707 KSSQKELEAMGLTPESEAIIEEAHPTDVSISKTALYSRIGTAETAEVEKPAQLLFQOPDLD 766
QY 752 DTGDEWIETPSQR---KWECDFVDKHQRTLEVRSAVELSKSAGLNQLLVLEDRARD 808
DB 767 SALQIARAEIITKEREVSSEWQXYESREVMEKIVAEYEKT-----IAQMIEDEQRE 821
QY 809 K-----VYMQQAIGD-----RLINDLQRFSSOKHAL---NRPVEFRQWVYESY 849
DB 822 KSVSHQTVQGLVLEKEQALADLNSVEKSLADLFRRYERKMEVLEGFGRKNEEVLKRCQAQY 881
QY 850 SSRATR 855
DB 882 LSRVKK 887
RESULT 12
ABJ25640
ID ABJ25640 standard; protein; 1555 AA.
XX AC ABJ25640;
XX DT 16-APR-2003 (first entry)
XX DE DE
XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX KW cancer; contamination; biofilm; antibody; immune response.
XX OS Aspergillus fumigatus.
XX FN WO200286090-A2.
XX PD 31-OCT-2002.
XX PF 23-APR-2002; 2002WO-US013142.
XX PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.

31-OCT-2002.

23-APR-2002; 2002WO-US013142.

23-APR-2001; 2001US-0285697P.

27-APR-2001; 2001US-0287066P.

05-JUN-2001; 2001US-0295890P.

09-JUL-2001; 2001US-0303899P.

31-AUG-2001; 2001US-0316362P.

(ELIT-) ELITRA PHARM INC.

Jiang B, Tishkoff D, Zamudio C, Ershkin AM, Hu W, Lemieux SM; WPI; 2003-093124/08.

New purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*, useful for treating or preventing infections by *A. fumigatus*, or for treating a non-infectious disease in a subject e.g. cancer.

Disclosure; Page; 175pp; English.

The invention relates to novel purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organism invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This sequence represents a protein of one of the essential genes of *Aspergillus fumigatus* of the invention

Sequence 1832 AA;

Query Match 2.4%; Score 176; DB 6; Length 1832;

Best Local Similarity 19.2%; Pred. No. 0.00037;

Matches 229; Conservative 144; Mismatches 458; Indels 362; Gaps 55;

9 RNSPVERII-----NRLNNDVNLGQVADTT-----LTPHRKELAEDEDFGRHDKI 57

332 KNSPSEEDLAARWQAEISDD-----AETMPTEDDLARWQAEISDDDDDDLLDDDT 393

58 YRALNFLYRKDDSLNQAENAFIEAKAASNNWVPKHAADFTLPLWSKEPPRAATA---- 113

384 TNA-----QRPPEAANDHNM--DTSMQLQSPFGTPENLARPKQPVSYTPHQPSSTLLS 436

114 ---GQWALQTVLEVLNRF--MPPN--NTPGTFG-RILSG-----PSGLSRPTST 158

437 GIPAQNTAAQNTNMSYSFAQAPPNPVTTAESEFAESKEGYSYDIPEDLARPRRA 496

159 NTKRKDEPANY--TFADPPKRLSTRSATGPPHIGAAIPLKFPDPVNTGSKRPSLESEN-- 215

497 VANSRTVVAQGTVPKPPRS--SSIPAPPLKASTVS--PAPLGTSSTAPTAPQKNFPE 551

216 -----NOCTKRAKGLSNVAAAAPPV-----PIASALDKVPTERRHANTDPTATG 262

552 ELPLPPRPKRSPASSGRYTPN-APVSAPSLPSIPIPPANQYNSVPGACQSNIGPDPPQ 610

263 HRRADQVDSFTSQGTSYGS-----SVFSACRINQSTTOSFFAPPSQPREKRPVATVF 317

611 LQOPERLDVSNLLAPNVPAPAVPSTASYSRPPGVQAGVKKPPSPRYSPAPPQSTNA 670

318 EAGHLIESPSKGRITKSHIDNQLSSSSQGETSFSIYSEFSSGGBGGAIPESRSNGLA 377

671 VAA-----APRNRYSQPASISGGGAA-----LQOPRTSSPLAYHEKIHVEDQG 715

378 RSEESARSQVQVHAPVVAARLRNIWPKFKWLHEAPLAVAWETVTRFLMHCKVLEBSLG 437

716 QSEB--RPQLQSTASP-----PPLNHSHP-----SEQPVSSBENKG 748

438 -----LKYPDPSNSTARDVTDINKLYRLDAFRGKPPPEKPPNDVFV-TAMTGNFESKGS 490

749 PSGVDVLENPFPSTR-----PQSPPKNPYAPSAITNEFANRVA 787

491 AVV-----LSAVLDYNPNSTAP-----LYLVKXKPLMFEQGCRLT 527

788 PVSTGPPIAGWTGVLSNSTEESPPVPRRQTSQSPQTLSPRLSVPSLDP--FORPASVH 845

528 RRFQDFRFFILLPSPTST---SPS-----VPPVVSQKPGCAVEEVIQWLTMGQHSLVGRQ 579

846 GSTSPTTRTVNAPYAPVPTNRPASQVLEFPITPDTGGQQLDSLE-----R 889

580 WRAFFAXDAGYKPLREF-----QLRAEDPKPIKE-----RVHFF--- 615

890 WKG-----APIFKFGGAVISCFKHIPYSAGQAAPMIKSCPGVRISQLNDWL 940

616 -ASTGIPTREDVFTKRSVPAEEVEQRTFKYSQMLDQLLDNNNTWQHLKFSRLQ 674

941 PAAGGIQVHQFGLPKGKS-----KKDLVAMLSKIAAFENANNIPDIDLSP 986

675 GLSKTYAINTLEPHQIRHHKTDLLSPSGTGEVMDGVGRMSRSVAKETRDVL----- 726

987 DASK-----LREEXT-LLMKVIRVLVENDGVLEGSVEAQSLRNLFPNLQDSG 1034

727 ---GLGDV-----PSAVQGR-----FGS-AKGMW-VIDVDDTGDDE 756

1035 PNQSLGVDFTSATLQPLNAPSQPDVDSRSDVLLRDTLVLGEREKAVAAVDRKLMGHA 1094

757 DWIETPYSORKE--CDFVDKQORTLEVRSAVELKSAGINLQLLP-VLEDEARDVKVM 812

1095 MIASRMDRSVWQVQVEFVR-----EVSASRSTSLAAFYELLAGNIEESIDELVPP 1149

813 RQAIGDRLINDLQRFSEQKHALNRPVEFQWVYESYSSRA----- 853

1150 SARAGLQWISKVDGH-GPAXNSLDGLDSWRETGVGLVLSNRSPPDQALVALGRLLLSVNR 1208

854 TRVSH-----GRVPFLAGLPDSQEBETLFLMNSGDPKCKYKQLQDIADWLQKRCDTLKS 908

1209 TEAAHICFILSRVAVFGGLDDPQAN-----IVLVGDHQRLLSSCAALYNDSDSILLTEAVEF 1264

909 KLINRVGRSA-----YIYMIADPWGLVEENEHVGFSSKFRDEESFTLLSDCDLVAR 962

1265 ATSVLAGSSVSTLPHLLAFKLIHAWSLAERGR-----KSEAQQY-----CDATAA- 1309

963 SPAHPFSDIQVRVAFKPELHSLDXDVIIFSTKGDVPLAKLSGGDGDGMW 1015

1310 -----ALKATTKEGSHYNOHLFF-----GVDELSARLRETTSDGGSWI 1348

RESULT 14

AB565616

ID AB565616 standard; protein; 2271 AA.

XX

AC AB565616;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 23640.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
DR N-PSDB; ABL09719.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 23640; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2271 AA;

Query Match 2.3%; Score 173.5; DB 4; Length 2271;
Best Local Similarity 18.7%; Pred. No. 0.00087;
Matches 191; Conservative 147; Mismatches 395; Indels 287; Gaps 44;

Qy 58 YRALNPLYWRKDSLNQAEANFFIEAKAASNVVKAHA-----DPDTLPMSKEPPRAA 111
Db 935 FELLNFIKLRKNDGDIRYQINVKIEWKVLAHNKMANSKTRIIIEPDCLRWS---PLLS 991
Qy 112 TAGQWALQTVLLEVLNRMPPNNPTGRTFGRTLGPGLSRPTSTNTYKRDE----- 165
Db 992 VS-----KLPNIIRKIPISNREYLSNQMETNSSFKEINEDKISVQGVSTSSKVNQESSELST 1047
Qy 166 ----PANVTTFADPKRSLTSATGPPHGAAILPKFPDPVNTGSKRPSLESENINQCTKR 221
Db 1048 KNRGPSSEKLIDREENQR-----AYPEFLKSASNEAKPKLTAPQPN-LEQ 1096
Qy 222 AKGLSDNVAAAAAPPV-----IASALDKVPTRRHRANTRDPTATGHRRAQVDSF 272
Db 1097 SFGELSDSDCKSSKDLSEQLTKRAQKLAQRNDLIP---HTNKR-----KHF 1140
Qy 273 DTSQGTSGSVSFSACHNOSTOSSPEAPPSPREKRPVDATVFEAGHLIESPS----- 327
Db 1141 D-----RPLENNTSTDHNPASQ-----ADTAGTENGKSBETAKENC 1181
Qy 328 KGRTTKSHINDQLSSSQGETSFSTYVESFGSGGEGALPEPSRNGLARGEESARSQV 397
Db 1182 KNVVEKEHVD-KPVN-----IMPKLRGK-----QSNQ-KRSEEPQFNFEN 1218
Qy 388 QVHAPVVAARLNINWPFPKWLHEAPLAVAVEVTRLFPMCKVDLEDESLG-----LKYDP 442
Db 1219 PTDSPVAVASLSP-----AKEITREPVKVA-----KTESPQKQESAKHDA 1259
Qy 443 SWSTARDVTDIWKTLRLDADFGRKPFPEKPPNDVFVTAMTGNPESKGSVAVLSAVLDYNP 502

Db 1260 EPNSVSDQV-----LDA-----VAKTKTKTLFSDVASYNTRD-----VKYNS 1296
Qy 503 DNSPTAPLYLVKLPLMFEQGCRLTRRFGDRPFELIIPSTSTSPSPVPVWVSKOPGAVE 562
Db 1297 EEKPMATAQEPKESPT--EAGIKETKP-----LEVEQAKAKAKSPKPAVAE----- 1342
Qy 563 EVIQWLTMGQHSVLGROWRAFFAKDAGYRKLRFQLEAEADPKPIIKERVHFFAETGTF 622
Db 1343 -----TVLQH-----VHAPVNTKEISTESK--LELDVHY-----LKL 1373
Qy 623 RPDVFKTRSVVPAEBPV-----EORTEFKVSOMLDMLQLDNNWTWQPHLKLFSRIQLG 675
Db 1374 SPDSALNPPVASVOKPVPPEPPAPKEKADLQAAQPSALQK-----QPEIEMVKTE-- 1425
Qy 676 LSKTYAINTLEPHQIRHKTDLLS-----PSGTGEVN-NDGVGRMSSSAKRIIRDV 725
Db 1426 -TKTETPVSNPP---SNRTDVSLSVEQPVKAAPEKVEQPMVKEAVDIKDKALKKPV--- 1477
Qy 726 LGLGDPVSAVGREGSAGKMWVIDVDDTGDDEWLETYPQSKWECDFVDKHORTLEVRVSV 785
Db 1478 ---PDVP-----VVRPEATSCCKIDHSKTVFHEKEAIKTDQQLVQVKE- 1519
Qy 786 ASELKSAGLNQLLPVLEDRARDKVRQAIGRLINDLQRFSEQKHALNRPVEFRQWV 845
Db 1520 -EEKSNARAPSAPIPI-----RDKIQLKGNEHAQLNSIPSQFPLNQMPNYHTSQWQWE 1573
Qy 846 YESYS-SRATRVSHGRVPPFLAGLPDSQEETLNFMLN-----SGFDPKKOKYL 891
Db 1574 YYGYNLSHLDASQKQKQFHKDLATTWAYTHNFQNLVQSANLAWQHAHQHQTKEKH 1633
Qy 892 QDIAWDLQKRKCDTLKSKLNRVGRSAVIYMIADFWGLBENEVHVGVFSSKFRDEESFT 951
Db 1634 VE-----RNSCKKSEQN-KVYSSVNVVSTA-----REDAHVQHCNEYAANNQAAL 1678
Qy 952 LLSC-----DVLVARPAHPFPDIQVRVAFVKDELHSLDKDVIIFSTKGDVPLAKL 1006
Db 1679 YNOKCASQKQKQKSLANVQNPVPRQSNASNESTVLMNPVVISQAGSVPAKQKVEHG 1738

RESULT 15
ABRS2986
ID ABR52986 standard; protein; 1226 AA.
XX
AC ABR52986;
XX
DT 20-JUN-2003 (first entry)
XX
DE Protein sequence #SEQ ID 837.
XX
KW Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS Saccharomyces cerevisiae.
XX
PN EP1258494-A1.
XX
PD 20-NOV-2002.
XX
PF 20-DEC-2001; 2001EP-00130253.
XX
PR 15-MAY-2001; 2001EP-00111774.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;
XX
DR WPI; 2003-250078/25.
DR N-PSDB; ACC61028.
XX
XX New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:46:22 ; Search time 26 Seconds
(without alignments)
2783.831 Million cell updates/sec

Title: US-09-913-878A-2

Perfect score: 7397

Sequence: 1 MNPITPRKRNSPVEEIIINRL.....YEVLGDDDFDGIGFTNGDYG 1402

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5S_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6S_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 297 | 4.0 | 1114 | 3 | US-08-811-583-2 |
| 2 | 151.5 | 2.0 | 218 | 3 | US-08-811-583-3 |
| 3 | 148.5 | 2.0 | 1457 | 4 | US-09-436-874-2 |
| 4 | 142.5 | 1.9 | 1457 | 4 | US-09-713-273A-18 |
| 5 | 142 | 1.9 | 2409 | 6 | 5180808-2 |
| 6 | 141.5 | 1.9 | 909 | 3 | US-09-425-383-2 |
| 7 | 139.5 | 1.9 | 1312 | 3 | US-09-041-886-19 |
| 8 | 139.5 | 1.9 | 1312 | 4 | US-09-648-281-2 |
| 9 | 139.5 | 1.9 | 1312 | 4 | US-09-707-919A-19 |
| 10 | 139.5 | 1.9 | 1312 | 4 | US-09-083-268-3 |
| 11 | 139.5 | 1.9 | 1384 | 3 | US-08-976-255-11 |
| 12 | 139 | 1.9 | 1528 | 4 | US-09-845-917A-3 |
| 13 | 139 | 1.9 | 1583 | 4 | US-09-845-917A-4 |
| 14 | 137 | 1.9 | 1768 | 4 | US-09-489-039A-11704 |
| 15 | 136.5 | 1.8 | 2972 | 4 | US-09-579-181-2 |
| 16 | 136.5 | 1.8 | 3118 | 4 | US-09-579-181-1 |
| 17 | 129.5 | 1.8 | 1317 | 3 | US-09-083-521-7 |
| 18 | 128 | 1.7 | 1093 | 4 | US-09-252-991A-21827 |
| 19 | 128 | 1.7 | 3969 | 3 | US-08-061-376-5 |
| 20 | 127.5 | 1.7 | 1191 | 4 | US-09-921-039A-2 |
| 21 | 127.5 | 1.7 | 1191 | 4 | US-09-921-039A-4 |
| 22 | 127 | 1.7 | 1341 | 4 | US-09-252-991A-26785 |
| 23 | 125.5 | 1.7 | 511 | 4 | US-09-252-991A-26078 |
| 24 | 125.5 | 1.7 | 3418 | 3 | US-08-639-501-2 |
| 25 | 125.5 | 1.7 | 3418 | 3 | US-09-044-946-2 |
| 26 | 125.5 | 1.7 | 3418 | 3 | US-08-755-587-44 |
| 27 | 125.5 | 1.7 | 3418 | 3 | US-09-044-908-2 |

| | | | | | | |
|----|-------|-----|------|---|----------------------|-------------------|
| 28 | 124.5 | 1.7 | 2188 | 4 | US-09-328-352-7763 | Sequence 7763, Ap |
| 29 | 124.5 | 1.7 | 3418 | 2 | US-08-603-753D-4 | Sequence 4, Appli |
| 30 | 124.5 | 1.7 | 3418 | 3 | US-09-098-753-4 | Sequence 4, Appli |
| 31 | 124.5 | 1.7 | 3418 | 3 | US-08-986-106-4 | Sequence 4, Appli |
| 32 | 124 | 1.7 | 1581 | 3 | US-09-110-517-2 | Sequence 2, Appli |
| 33 | 122.5 | 1.7 | 801 | 4 | US-09-252-991A-27870 | Sequence 27870, A |
| 34 | 122.5 | 1.7 | 1742 | 4 | US-09-386-962C-4 | Sequence 4, Appli |
| 35 | 122 | 1.6 | 3898 | 3 | US-08-750-717-2 | Sequence 2, Appli |
| 36 | 121.5 | 1.6 | 582 | 4 | US-09-450-072-79 | Sequence 79, Appl |
| 37 | 121.5 | 1.6 | 582 | 4 | US-09-351-348-79 | Sequence 4, Appli |
| 38 | 121 | 1.6 | 521 | 2 | US-08-682-847-4 | Sequence 4, Appli |
| 39 | 121 | 1.6 | 1394 | 4 | US-09-845-917A-57 | Sequence 57, Appl |
| 40 | 121 | 1.6 | 1449 | 4 | US-09-845-917A-58 | Sequence 58, Appl |
| 41 | 120 | 1.6 | 889 | 4 | US-09-252-991A-18702 | Sequence 18702, A |
| 42 | 119.5 | 1.6 | 525 | 4 | US-10-132-920B-27 | Sequence 27, Appl |
| 43 | 119.5 | 1.6 | 1130 | 4 | US-09-442-100-4 | Sequence 4, Appli |
| 44 | 119.5 | 1.6 | 1130 | 4 | US-09-233-857-3 | Sequence 3, Appli |
| 45 | 119.5 | 1.6 | 1130 | 4 | US-08-939-106-4 | Sequence 4, Appli |

ALIGNMENTS

RESULT 1
US-08-811-583-2
; Sequence 2, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenegger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1114 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-811-583-2

Query Match 4.0%; Score 297; DB 3; Length 1114;
Best Local Similarity 20.4%; Pred. No. 2.5e-16;
Matches 175; Conservative 133; Mismatches 323; Indels 228; Gaps 32;
QY 562 EEVIOIWTGQSHLVGROWRAFFAKDAGYRKLRSFOLRAEDPKPIIKERVHFFAETGIT 621
|||

Db 386 EDIDNPLAV---SFVDBEWKLYSTDL-----LPK-----ASTGSG 418
Qy 622 FRPDVFK-----TRSVVPAEBFVQTEFKVQMLDMLQDNNW---OPHLKL--- 668
Db 419 VRTNIYERILSTRKGFVIGDKKFE-FLAFSSQL-----RDSVMVFAFRPGLTANDI 471
Qy 669 -----FSRI-----QLGSKYVAJMTLEPHOIRHKTDL---SPSGTGVWMDGV 711
Db 472 RAMWGDPSOIKNVAKYAALGOSFSRETL---SVLRHEIEVIPDVVKVGTSTVFSNGI 528
Qy 712 GRMSRSVAKIRIDVILGLGVPSAVQGRFSGAKGMWVIDVDTGTDBEDTETVPSORKMECD 771
Db 529 GKISGDFAHRAVSKGGLQVTPSAFOIRYGGYGVGVDFDSMK-----LSLRKSKSYKYESD 585
Qy 772 FVDEKQRTLEVSASSELKSAAGLNLQLPVEDRDARDVKRQAIGDLRLINDLQORPSEQ 831
Db 586 NI-----KLDVLG-WSKQPCVYLNQLITLS-----TLGVKDEVLEQKQKEAVDQ 630
Qy 832 KHALNRPVEFRQWVYESYSSRATRVSHGRVPPFLAGLPDSQBEETNLFMNSGDFPKKQYL 891
Db 631 LDAILHLSLKAQEALELMS-----PGENTNILKAMLCNGYKPDAPFPL 673
Qy 892 QDIANDLQKCKDTLKSCLNIRVGRSAIYMIADFWGLVENEVHVGS-----SKEDDEE 947
Db 674 SYMLOTFRASKLULDRTRSIFIPNRTVMGCLDESRTLEYGVQVFTGAGHGEFSDDL 733
Qy 948 ESF-----TLSDCDVVLVARSAPHFPSDIQVRVAVFKPELHSLKDVIIIFSTKGDVPL 999
Db 734 HPFNNSRSTNSFILKGNVVAKNPCLHPGDIRVLKAVNVRALHHWDCVVPQKGEKH 793
Qy 1000 AKXLSGDDYGDMAVWCVDPDEIVDFVNAEMPLEDLSRYLKDKXTTFKQLMASHGTGSA 1059
Db 794 PNECSGDLDDGIYFVCWDQDMIPPRQVQPMYEPAPSIQDHD----- 837
Qy 1060 AZEQTYDMIKSFHALQPNFLMCTN-----YKERLCVINNSVSNKPAIILSLVGNLV 1115
Db 838 -----VTIEVEEYFTNYIVNDSGLIIANAHVVPADR---EPDMAMSDPCKKLAEFSIAV 890
Qy 1116 DQSKQIVNEASWALQRLGGLSALPDPMYKSDSLWGRGEPHIIIDYKFSIARPAI 1175
Db 891 DFPKTVGPAEIPS--QLRPK-----EYDFDMKPD-----KTSYISERVIGKLPKVK 936
Qy 1176 KLELBAPHNAKAKDTEGAHFWDPLASVYTFPEISDKSRSSALLFTLKNRIGRV- 1234
Db 937 DKA-----PQASSIATFRDVARSYDADMEVDGFEDYIDEAF 974
Qy 1235 --EKEYGLRVKNEWRDSKDPYPRVNOYVEKWCATITPEAMDKSANYD-----SKV 1284
Db 975 DYKTEY-----DNKLGNLMDYYGIKTE-----ABILSGIMKASKTFDRKDAEASVA 1023
Qy 1285 IRLLELSFLADREMTW-----ALLRSTAPKLYYHKS-----PKFV 1321
Db 1024 VRALR-----KEARAFKRRNDIDMLPKASANTHYTHPTYWCYNQGLKRAHFISFP 1077
Qy 1322 WQWAGRLAYIKAOQTSRP 1340
Db 1078 WCYVDQLIQIKDKARNRP 1096

RESULT 2
US-08-811-583-3
; Sequence 3, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenecker, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sadger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-811-583-3

Query Match 2.0%; Score 151.5; DB 3; Length 218;
Best Local Similarity 24.9%; Pred. No. 7.2e-05;
Matches 52; Conservative 34; Mismatches 84; Indels 39; Gaps 5;
Qy 930 LEENEVHVGVGS-----SKFRDEESF-----TLSDCDVVLVARSAPHFPSDIQVRVAV 977
Db 13 LEYGVQVFTGAGHGEFSDLLHPFNNSRSTNSFILKGNVVAKNPCLHPGDIRVLKAV 72
Qy 978 FKPELHSLKDVIIIFSTKGDVPLAKLSGGDYGDMAVWCVDPDEIVDFVNAEMPLEPDLS 1037
Db 73 NVREALHHWDCVVPQKGRPHFNECSGDLDDGIYFVCWDQDMIPPRQVQPMYEPAPPS 132
Qy 1038 RYLKDKXTTFKQLMASHGTGSAKEQTYDMIKGFHFPALQFNFLMCTN-----YKERLC 1093
Db 133 IQLDHD-----VTIEVEEYFTNYIVNDSGLIIANAHVVPADR--- 170
Qy 1094 YINNSVSNKPAIILSLVGNLVDSQKGI 1122
Db 171 -EPDMAMSDPCKKLAEFSIAVDPPKTV 198

RESULT 3
US-09-436-874-2
; Sequence 2, Application US/09436874
; Patent No. 6521816
; GENERAL INFORMATION:
; APPLICANT: Froberg, Claus
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES FROM RICE AND THEIR USE FOR THE
; TITLE OF INVENTION: PRODUCTION OF MODIFIED STARCH
; FILE REFERENCE: GFB-10
; CURRENT APPLICATION NUMBER: US/09/436,874
; CURRENT FILING DATE: 1999-11-09
; EARLIER APPLICATION NUMBER: US 60/107,883
; EARLIER FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1457
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-436-874-2

Query Match 2.0%; Score 148.5; DB 4; Length 1457;
Best Local Similarity 18.0%; Pred. No. 0.004;
Matches 267; Conservative 213; Mismatches 532; Indels 471; Gaps 69;

QY 95 HADPOTLWSEKPPRAAATAGQWALQTVLLEVRNFMPPNPTGRTGLSPGSLR 154
DB 21 HARPASPPALLPPAALRRGRRLPAATTTTAVRSRLAP-----RAIAASTGRAS 71

QY 155 P--TSNTYRKDEPANVTADPPKSL-----TRSATGPIHGAIPKFPD----- 199
DB 72 POLVGRFTLDANSELKVTINPAQGSVAEINLEATNTSGSLILHWGALR---PDGFWLL 128

QY 200 PNTGSKPSPLESENLNCTKAKGLKLDNVAAPPPVPIASAL-----DKVPTRRHAN 254
DB 129 P-----SRPDGTVVYKRALRTPFTKSGDNLKLEIDDPVQAIEFLIPDEARNWYKN 184

QY 255 TRDPTATGHRRAQVDSFD--TSQTSYSGS-----VFSACRNQSTQSSFEAPP 303
DB 185 -----NQGNFOIQASQYQCGQSTATSTSTVVPDELVQISYLRWERKQKQ---YTP 235

QY 304 SOPREKRPVDATVFAG--HLIESPSKG-----RTTKSHIDNQPLSSSSQGETSFST 353
DB 236 EQEKEE-----YEAATELIEELKNGVSLKULRAKUTK----- 269

QY 354 YYESFPSSGGEGAIPERSNGLARSEARSQVQVHAPVVAARLNIWPKF--PKWLHE 411
DB 270 -----PEATDSNAPA--SESTVTKVPEELVQVQAYR--WEKAGKPNYAPE 312

QY 412 APLAVANEVTRLFMECKVLEDESIGLXVDSWSTARDVTDINKLYLXLDFAFRKPEPEK 471
DB 313 KQL-----VEFSEARKELOSELDKGT'S-----VEOLRNK----- 341

QY 472 PPNDFVFTAMTGNFSGSAVVLAVDYNPDNSPTAPLYLVKLPLMFEGQCLTRRF 531
DB 342 -----ILKGNIEKYSK-----QLKDKKYFSVERIQK-- 369

QY 532 PDRFEILLPSTSTSPSPVVPVQKQAVBEVQWLTMGQHSVGRQWRAFFAKDAGYR 591
DB 370 -KRDIVQLLKGKPTWBAQAEPTKQPTVLD-----LFTK----- 403

QY 592 KPLREFQRAEDPKPIIKERVHFFAE---TGITPRDVFKTRSVVPAEPEQOEKFEKVS 648
DB 404 -----SLQEQNCEVLSKLFKFGCKEILGIT-----TVALGKTKVHLATNWE 448

QY 649 QMLDMLLDLNNWQ--PHLKL-----FSRIOL-GLSKTYAIAIMTLEPHOIR 691
DB 449 LILHWALSKENGWQAPPSSILPSSGSLDKACETSFSEYELNGLH--CQVVEIELDDG 506

QY 692 HKHTDLLSPSGTGEYMDGVG---RMSRSVAKRIRDLGLGVPSAVQGRFGSAGKMWVI 748
DB 507 YKRMPEVLRSGETWNNKNSPYLDFSTKVANKYOT---GDA-----GK-GTAKAL--- 554

QY 749 DVDDTGDEDIWITPSQRKWCDFVDKHQRTLEVR--SVASELKSAGLNLQLLPVLEDRAR 807
DB 555 -----LERI-----ADLEDAQRLMHRFNIAAD-----LVDAQAR 584

QY 808 DK-----VKMROAIGRLINDIQRQFSEQKHALNRPVPRQWVYESYSSRA 853
DB 585 DNGLLIGITGIFWIFEMATROLIMKNVYKPREISKAQDRFTDLENNYTPQY---- 640

QY 854 TRVSHRVPFLAGLPDSQETLNFILMNS---GFDPKQKQLQDIADWLQK----- 900
DB 641 -----QELLRWMSAVGREGGEGVQQRIRDEILVIRNNDCKGGMWE 682

QY 901 ---RKCDTLKSLNTRVRSAYIMADF-----WGVLENEV-----HVGFSK 942
DB 683 EWHQKLNHTSPDDVVICALLDYIKTSDFDIGVYMDTLCKDGGITKERLLSYDRPIHSEPN 742

QY 943 FRDEEESFTLLSDCDVLVARGPAHPSPDIQVRVAV---FKPELHSLKDVIIFS*KGDPVPL 999
DB 743 FRSEKQDGLRLDGNMYSKAVHSGADLESAIATCMGYKSEGG-----FNVGVQINP 796

QY 1000 AKKLSGG-----DYDGDGMWVWCPDPEIVDGFVNAEMPLEDPLSRYLKOKITFKQL 1050

797 VKGLPSGFKPLLEFVLDHVEDKS-----AEPLLEGLEAPAEHLPLLLSGSPERMKDLIFLD 852
1051 MASHGTSGAASQTTYDMI-----OKSFHF-ALQPNFLGMCNTYKERLCY----- 1094
853 IALDSTFRTAVER-SYEELNNVEPEKIMYFISLVLENLALSTDDNEDILYCLKGMNQALE 911
1095 INNSVSKKAILLSSLVG-----NLVDOSKO--GIVFNEASWQ--LRREL 1136
912 MAKQKNQWALYAKAPLDRTRIALALASKGEQYVNLMPQSAEYLGSLINIQWAVNITFEI 971
1137 LGGA-----LSLPDPMYKSDSMIGRGE---PTHIIDYL-----KFSIARPAIDKELE 1180
972 IRGGSAAATLSALLNRIDPVLNRVAQLGSWQV-SPVEVSGYIVVDELLAVQNKSYDKPTI 1031
1181 AFNANAKAARDTEG-AHFWDPDOLASYYTFFKEISDKSRSSALLFTTL--KXRIGEVEKE 1237
1032 LVAKSVKBEETPDGVGVVITPDMED---VLGSHSVRARCNKVLFACTPDPNTLSLQGH 1088
1238 YGRLVKXNKMRDSKDPYVVRVNOVYKWCATPEAMDKSGANYDSKVIIRLLELSFLADRE 1297
1089 DGKVFSEK-----PTSADITYRE---IPSELSQSGS-----LNAEAGQA 1124
1298 MNTWALLRASTAPKLYYHKSPKPVWQAG---RQLAYIKAOMTSRPGEAPALMTAFMYA 1354
1125 VPSVSLVKKKFLGK-YAISAEBSFSEEMVGAQRNVAYLKGVPSWVGVTVAIPEGTPE 1183
1355 GLMPDK--KFTQYVARLEGDSGEVDPPEVYVLDGDDDFDGIG 1395
1184 KVLSDSEINVEAQTIQMLKKG-----LAQDDFSALG 1214

RESULT 4
US-09-713-273A-18
; Sequence 18, Application US/09713273A
; Patent No. 6620987
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; TITLE OF INVENTION: STARCH R1 PHOSPHORYLATION PROTEINS
; FILE REFERENCE: BB1158 US CIP
; CURRENT APPLICATION NUMBER: US/09/713,273A
; CURRENT FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: 60/081,143
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: PCT/US99/07639
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/679,933
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 1457
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-713-273A-18

Query Match 1.9%; Score 142.5; DB 4; Length 1457;
Best Local Similarity 18.1%; Pred. No. 0.013;
Matches 268; Conservative 212; Mismatches 532; Indels 471; Gaps 70;

QY 95 HADPOTLWSEKPPRAAATAGQWALQTVLLEVRNFMPPNPTGRTGLSPGSLR 154
DB 21 HARPASPPALLPPAALRRGRRLPAATTTTAVRSRLAP-----RAIAASTGRAS 71

QY 155 P--TSNTYRKDEPANVTADPPKSL-----TRSATGPIHGAIPKFPD----- 199
DB 72 POLVGRFTLDANSELKVTINPAQGSVAEINLEATNTSGSLILHWGALR---PDGFWLL 128

QY 200 PNTGSKPSPLESENLNCTKAKGLKLDNVAAPPPVPIASAL-----DKVPTRRHAN 254
DB 129 P-----SRPDGTVVYKRALRTPFTKSGDNLKLEIDDPVQAIEFLIPDEARNWYKN 184

QY 255 TRDPTATGHRADQVDSFD-TSQGTSYGSS-----VFSACRHNQSTQSSFEAPP 303
Db 185 -----NGNQFQIQLOASQYQCGQSTATSSVVPEDLVQIOSYLRWERKGQS---YTP 235
QY 304 SQPREKRPVDAVTFEAG--HLIESRSG--RTTKSHIDNQPLSSSSQGETSPST 353
Db 236 EQFKEB-----YEAARTELIELNKGSLSKLAKLTKT----- 269
QY 354 YYESFSSGEGAIPEPSSNGLARSEARQOVQVHAPVVAARLNIPKPF--PKWLHE 411
Db 270 -----PRATOSNAPA--SESIVTKVPEELVQVQAYIR--WEKAGKNYAPE 312
QY 412 APLANAVEVTRLFPMCKVLEDESLGLKYDPDSNSTARDVTDIWKTYLRDFAFRGPFPEK 471
Db 313 KQL-----VEFEARKELOSELQDGT-----VEQLRNK----- 341
QY 472 PPNDVFVTAMTNFSGAGSVLVAVDYNDPNSPTAPLYLVKLKPLMFEQOQCRLTRFG 531
Db 342 -----ILKNIEFKVSK-----QLDKKYPVSVERIQK-- 369
QY 532 PRFFELIPSTSTSPSPVVPVVKQPGAVEVIOHWTGQSHLVGRQWRFAFFAKDAGR 591
Db 370 -XRDIVQLLKXKHPVNEAQVETPKQPTVLD-----LFTK----- 403
QY 592 KPLREELQAEADPKPIKERVHFAE---TGITFRPDVFKTSVVPABEVEQRTFEKVS 648
Db 404 -----SLOEQDNCEVLSRFLFKFGDKELGIT-----TVALGKTKVHLAINYEP 448
QY 649 QMLDWLLQLDNNWQ--PHLKL-----FSRIQL-GLSKTYAINTLEPHQIR 691
Db 449 LILHVALSKENGWQAPPSSILPSGSSLLDKACETSFSEYELNGLH--CQVVEIELDDGG 506
QY 692 HHKTDLLSGTCEVNDGVG---RMSRSVAKRIEDVLGLGVPSAVQGRFGSGAMVVI 748
Db 507 YKRMPEVLSRGETWKKNGSGDFYLDFTKVAKNTKDT---GDA-----GK-GTAEAL--- 554
QY 749 DVDDTGDDEWIEPTPSQRKWECDVFDKHQRTLEVR--SVASELSKAGLNLQLLEVEDRAR 807
Db 555 -----LENI-----ADLEDAQRLMHRFNIAAD-----LVDQAR 584
QY 808 DK-----VKMROAIGDLINDLQROSEQXHALNRPVEFRQWYESSRA 853
Db 585 DNLGLGIIGFVWIGFMATRLQIWNKNYNKPREISKADQRTDLDLNNRYTPOY--- 640
QY 854 TRVSHGRVPLAGLPSQETLFLMNS---QFDPKKQYLOQIAWDLQK----- 900
Db 641 -----QELLRWMSAVRGEGDVQORIRDEILVIQRNNDCKGMWE 682
QY 901 ---RKDTLKSUNIRVGRSAYIYMTADP-----NGVLENEV-----HYGFASK 942
Db 683 EWHQKLHNNTSPDDVVICQALLDYIKSDFTGYNDTLKKGITKERRLLSYDRPIHSEPN 742
QY 943 FRDEESFTLLSDCDVLVARSAPHFPSDIQVRVAV---FKPELHSLKDVIIIFSTKGDVPL 999
Db 743 FRSEQKSLRLDIGNYMSRLKAVHSGADLESALATCMGYKSGEG-----FMVGQINP 796
QY 1000 AKKLSGG-----DYGDMAWCWDPEIVDGVFNAMPELPPDLSRYLKDKTKTFKQL 1050
Db 797 VKGLPSGFPKLLFILDHVEDKS-----ARPLLGLLEARAEHLPLLGSPERMKOLIFLD 852
QY 1051 MASHGTGSAKEQTYDMLI-----QKSFFH-ALQPNFLGMCNTYKRELCTYI-----NNSV- 1099
Db 853 IALDSFTFRATER-SHEELNNVEPEKIMYFISLVLENLALSTDDNEDILYCLKGNQVAE 911
QY 1100 -----SNKPAILISSLVG-----NLVDQSKQ--GIVNEASWAQ--LRREL 1136
Db 912 MAXQKNQWALYAKAFDLRTRFALASKGQYQYVLMQPSAEYLGSLINIDQWAVNFTBEI 971
QY 1137 LGGA-----LSLDPMYKSDSWLGRGE---PTHIDYL-----KFSIARPAIDKELE 1180
Db 972 IRGSAATLALLNRIDPVLNRVAQLSGMQVISFVEVSGYIVVVDLLAVQNKSDKPTI 1031
QY 1181 AFHNAMKAADTGDG-AHFWDPDPLSAYTTPFKBISDKSRSSALLFTTL--KNRIGEVEKE 1237

RESULT 5
5180808-2

Patent No. 5180808

APPLICANT: RUOSLAHTI, ERKKI I.

TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
ANTIBODIES, AND METHODS OF DETECTING THE SAME

NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/441,179

FILING DATE: 27-NOV-1989

SEQ ID NO:2:

LENGTH: 2409

5180808-2

Query Match 1.9%; Score 142; DB 6; Length 2409;

Best Local Similarity 18.0%; Pred. No. 0.037;

Matches 254; Conservative 162; Mismatches 509; Indels 490; Gaps 66;

QY 109 RAATAGQWALQTVL-LEVLNRFPNPNPTGRTFORTLSGPGSLGRPSTNTKRDPEA 167
Db 312 RAQCGGLLVRTLYRFENQTFPPDPDSREDAYCFKRMSDLVIGHIDSESK-EDEPC 370QY 168 NVTFADPPKSLRSATGPPHIG--AAIPLKFPDPVNTGSKRPSLESLNQCCKRAKG 225
Db 371 -----SEETDVFHDLMAELLPEFFDIIIDLYHSENEEEBEECAN----- 411QY 226 LSONVAAAAAPPVPIASALDKVPTRRHAN-----TRDPTATGHRADQVDSFDTSQ 277
Db 412 -----ATDVTTPSVQVINGKHLVTPKDPPEAAEARRG-QFESVAPSQN 455QY 278 TSYGS-----SVFSACRHNQST--TQSSFEA---PPSQPREKR-----PVDA 314
Db 456 PSDSSESDTHPVIAKTELSTAVQPNHSTETTESLSTVTHKPTYPETSEHFSGEEDVFP 515QY 315 TV-----FEAGHLIESPSKGRITTKSHIDNQPLSSSQGETSFSTYYIESPSSGGEAIP 369
Db 516 TVPFHEEFESG---TAKGAESVTERDTEVGHQAHEHTEPVSLEFPE---ESSGEAIDQ 568QY 370 PSRNGSLARS-----BESARSQVQVHAPVV-----AARLNINWP-----K 404
Db 569 ESQKIAPARATEVTFGEVEKSTSVTYTPIVFPSSASAYVSEEAVALIGNPWPDDLST 628QY 405 FPKWLHEAPLAVANEVTRLFPMCKVDL-----EDSLGLKYDPDSNSTARDVTDIW 454
Db 629 KESVVEATPR---QVVLESGSSIPITEGSGEAEDEDTMTVMTDLS---QRNTD-- 679QY 455 KTLVRLDAFR--GKPFPEKPNVDFVTAMTNGFESKSAVVLVAVDYNDPNSPTAPLYL 512
Db 680 -TLITLSDTSIIIESFPFVETATTIYPVS-----EQPSAKVPTKVFSETDSEWISSIT 732QY 513 VKLKPLMFEQOQCRLTRFRGDPFRFEILIPSTSPS--VPFVVSQKPGAVEEVIQWLTWG 571
Db 733 VEEKRKEEGSTT-----GTASTFEVY--SSTQKSDQLILPFLESESNVAT----- 776QY 572 QHSLVGRQWRAFFAKDAGYKPL-----REFQLRAEDPKFIKER---VHFAET--- 618
Db 572 QHSLVGRQWRAFFAKDAGYKPL-----REFQLRAEDPKFIKER---VHFAET--- 618

Db 777 -----SSDSGTRKSFMSLTPQTOSEREMTDSTPVFTETNLENLGAQTTHS 823
Qy 619 -----GIFRFP-----DUFKTSVVAPEBPVQRTFEKVSQMLDQLDN---NTW 662
Db 824 SIHQGVQEGLTILRSPASVFMWEGSGEAAADPETTTVSSFLNVVAIAQAEVACTL 883
Qy 663 QPHLKLFRIOQLGSLKTYAINTLEPHQIRHHKTDLLSPSGTGEVNDGVRMSRVAKRI 722
Db 884 SPHVE-----TTFSTEP-----TGLVLTVMVDRVVAENITQTSREIV--I 921
Qy 723 RDLVLGLGVPSAVQGRFSGAKMGWIDVDVDDTCDEDEMIETYSQRKWCDFVDKHQRTLEV 782
Db 922 SERLGEPNYGASIRG-----FSTG-----FPLEEDFGDF-----REY 954
Qy 783 RSVASELSAGLNLQLLPVLEDRARDKVMRQAIQDRLLNDLQ-----ROFSE 830
Db 955 STVSHPI-----AKEETVMWEGSGDAAFRDTQTSPSTVPSVHLSHSD 998
Qy 831 QKHALNRPVFRQWVYESYSSPA-----TRVSHGRVPFLAGLPDSOEFTLFLMNSGF 883
Db 999 SEGPSSTWVSTSAFFWEEFTSSAEGSGEQLVTVSSVVVPLPSAVQKFSGTASSIIDGL 1058
Qy 884 -----DPKQKYLQDAWDLQKKKCDTLKSLNIRVGRSAYIYMI---ADFWGVLEEN 933
Db 1059 GEVGTVNEIDRSTILPTA-EVEGTAKPVEKEEVKVSCTVSTNPTQTIETPAKLHRSQEVN 1117
Qy 934 EVHVGFSSKFRDEE-----SFTLLSDCDVLVARSAPHPPSDIQV-----RAVFKPELHSL 985
Db 1118 PYRQIESETTEGEOIQEBSKSP-----ESPONSATEQTIQDSQTTTETELKT- 1165
Qy 986 KDVIITFSTKGDVPLAKLGGGDDYDGMWVCHWDPEIVDGFVNAEMPLEPD--LSRYLKK 1042
Db 1166 TYSVLVTKYTSDDKEKEED-----TSLVNMSTP-DPDANGLESTTL 1209
Qy 1043 DKTTFK-----OLWASHGTGSA--AKEQTYDMMIQKSFHFALQPNFLGMCTNY 1088
Db 1210 PEATEKSHFFLATALVTSIPAETHVVDTSPIKKEEST-----KHPPKGMRT- 1256
Qy 1089 KERLCYINNSVKNPAIILSSVLGVLVDQSKOGIVFNEASWALRELLGALSPLDPMY 1148
Db 1257 -----IQESDTELLF----- 1266
Qy 1149 KSDSMLGRGE-----PTHIIDYKFSIARPAIDKELEBAPHNAMAKAANDTEDGAHFWDD 1202
Db 1267 -----SGLSGSEEVPLTPTESVNP-----TEVEQINNTL----- 1296
Qy 1203 LASVYTFPEKESDKRSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSK- 1252
Db 1297 -----YPHTSQVESTSSDKIEDF-----NRMENVAKEVGLVQSOTDIFEGSGSVTTLIEI 1348
Qy 1253 -----PYPRV-----NOVYEKWC-----AITPEAMDKSGANYDSKVRILLEL 1290
Db 1349 LSDTGAEGPTVAPLPFSTDIGHQPQNTV-RWABEIQTSRPQTIETQDSNKNSTAEINET 1407
Qy 1291 S-----FLADR---BMTWALLRASTAPKLYYHKS 1317
Db 1408 ITSTSDFLARAYGFEMAEFVTSAPKPSDLYEPS 1442

RESULT 6

US-09-425-383-2
; Sequence 2, Application US/09425383
; Patent No. 6194637
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Shi, Jinrui
; TITLE OF INVENTION: Maize DNA Ligase I Orthologue and Uses
; FILE OF INVENTION: Thereof
; FILE REFERENCE: 0962
; CURRENT APPLICATION NUMBER: US/09/425,383
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/108,793
; EARLIER FILING DATE: 1998-11-17

; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Zea mays
; US-09-425-383-2

Query Match 1.9%; Score 141.5; DB 3; Length 909;
Best Local Similarity 19.6%; Pred. No. 0.0071;
Matches 197; Conservative 138; Mismatches 361; Indels 311; Gaps 48;

Qy 152 LSPTSTNT-----KRDPEANVTADPP-----KRSLTRSATGPIHGAALP 194
Db 4 LTATTTTATTSILLRCPLPSPKVTTLATPAILFCIRSRFSASASASVPSQRA 63
Qy 195 LKFPDPVNTGSKPSLESENLCQTKRAKGLSDNVAAAAAPPVPTASALDKVPTERHAN 254
Db 64 TAAAAACKREKP-----PTARAKADAVPAAAAPOMS-ASGGSGAGKXTVAD 111
Qy 255 T--RDPATATGHRRAQVDSFTSQGTSYGSVFSACRHNQSTTQSFEEAPPQPREKR-- 310
Db 112 VLMGNARAAAASAKKAAPSPKPKQAKTDGAEV-----EPEAAAVAEKPSPSPARSKAS 164
Qy 311 -PVDA-TVFEAGHLIESPSKGRTRKSHID-----NQLSSSSQOGETSFSYVESFPSSGGE 364
Db 165 SPAKSKLLPDSSTSAKSGPAALQAQDAKENPP---SPKRSKTLAKSQTQPS--GE 219
Qy 365 GAI-----PEPFSRNGLA-RSEESARSQOVHAPVVAARLNIPKPPKWLHEA 412
Db 220 GVVLGHTDGKSPSPTKAKGQASQSEKKRAASPKAKTTDS-----PKTEK----- 268
Qy 413 PLAVANEVTRLFMHCKVDLEDSGLKYDPSWSTARDVTDIWKTLRLDAFRGKPPPEK 472
Db 269 -----NITLKKKGEFDPM-----AVAYWK-----PGBP 294
Qy 473 PNDVFVTAMTGNPESKGSVAVLSVLDVNDPNSPTAPLYLVKLKPLMFQCGCLTRRFQ 532
Db 295 VPFLFLARALDLISNBSGRIVITEIL-----SNVFTVMATP 332
Qy 533 DRFEELIIPSTSTSPVPSVSKPGAVEEVIQWLTMOHSLVGRQWTAFAKQAGYRK 592
Db 333 D---DLLATVYLSANRIAPP-----HEGIE-LIGDASVIRALAEAYGRKEEHVK 379
Qy 593 PLREF-----QLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVPAEEV 639
Db 380 DLKELGDLGLVAKASRSSQKMKPKPIIARVL-----STFR-----TAKESG 424
Qy 640 EQRTFKVQMLDMLQLDNNNTWQPHLKL-FRIQLGLSKTYAINTLEPHQIRHHKTDLL 698
Db 425 KDSQDKRNHMKGLLVAATDCEPQYITRLQSKMRIGLAETVQVAGLQAAVYADKNPL- 483
Qy 699 SPSTGEVNDGVGRMSRVAKRIKRDVLGLD--VPSAVQGRFGSAGKGMVVI----- 748
Db 484 -PPVQSPFEE-----AAKIIQAVSVLYIKVIPAILE-----VGWKLPECTCKFI 531
Qy 749 -----DVDDTGDDEMIETYSQRKWCDFVDKH---QRTLEVRVASE----- 798
Db 532 GVPVGMPLAKATKSVSEIIDKFCQLEYTCYKVDGERAQIHCWEDGSGVEIYSRNERNTG 591
Qy 789 -----LKSAGINLQLLPVLEDRARDKVMRQAIQDR-----LINDLQ 827
Db 592 KYPDVVDAVSRFRKATVKSFLVDCEI--VAYDREKOKILPQLLSTRAKKGVTTINDIKVS 649
Qy 828 -----FSEQKHALNRPVEF-RQWVYESYSSRATRVSHGRVFPFAGL-----PDSGE 872
Db 650 VCTFGFDILYNGKPLQLQOLKIRREHLNYSFE-----EVPQVFKLATAISNDLE 700
Qy 873 ETLNFL---MNSGFDPKQKYL-QDIAWDLQKCKDTLKSCLNI--RVGRSAYIYIADF 926
Db 701 ETQKFLDTAVNSCEGLIITKDQATYEPKASNNWMLKKDYMDSIGDSLVLPIAF 760
Qy 927 WG-----VLEENE-----VHVFGSKFRDEEESFTLLSDCDVLVARSF 964

[illegible]

RESULT 7
US-09-041-886-19
; Sequence 19, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabilzaden, Sharoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041.886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-041-886-19

```

Query Match      1.9%; Score 139.5; DB 3; Length 1312;
Best Local Similarity 17.6%; Pred. No. 0.02;
Matches 145; Conservative 98; Mismatches 265; Indels 315; Gaps 34

QY      10  NSPVBEIINRLNDY-----NLGLQCVADTTLTPH-----RR 41
Db      381  NEELEALELNDVNGCHDPDMRYNEENYGVVSTYDSSLSSYTVPLERDENSEELKEEAA 440
QY      42  KELAESDEFGRHDKIYIPALNPLYWRKDDSLNQBAENFFTEAKAA-----SSNVV 91
Db      441  NQLAAEIESSAQ---YKARVAL-----ENDDESEEKTVAVORNSSEEGHSINTRENKYI 493
QY      92  PRAHADPDTLPKS---KEPPRAATAG-----QQWALQTVLLEVL 127
Db      494  PPQQRNREVISMGSGRQNSPRMGQPGSGSPMRSTSHTSDFNPNSGSDQ-----RVV 545
QY      128  NREMPDPNNTP---GRTFGRTLSGSGSLRSTSTNTKEOEPAENVTFADDPKESLTSRAT 184
Db      546  NGGVNPFSPCPSPSRGPRSPRVQSGNSLPPRAATPTR---PPSR-----PPSRP-SRPS 596
QY      185  GPPFHGAAILPLFPDPVNTGSKR-----PSELESENINQCTKRAKOKLSDNV----- 230
Db      597  HPSAHSQS-----PAPVSTMPKRMSEGGPRMSPKQORHPNKHVSAGRGSISGSEFVS 650

```

[illegible]

```

RESULT 8
US-09-648-281-2
; Sequence 2, Application US/09648281
; Patent No. 6515197
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Transgenic Animal Model of
; TITLE OF INVENTION: Neurodegenerative Disease and Methods of Use
; FILE REFERENCE: P-CE 4336
; CURRENT APPLICATION NUMBER: US/09/648,281
; CURRENT FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-648-281-2

```

| | | | | |
|-----------------------|------------------|--|-------------|--------------|
| Query Match | 1.9%; | Score 139.5; | DB 4; | Length 1312; |
| Best Local Similarity | 17.6%; | Pred. No. 0.02; | | |
| Matches 145; | Conservative 98; | Mismatches 265; | Indels 315; | Gaps 34; |
| QY | 10 | NSPVEIIRNLNDY-----NLGQCQVADTTLTPH-----RR 41 | | |
| Db | 381 | NELEALENDVNGWDNDMFRNEENYGVSIVDSSLSSTVPLERDENSEEFLKREARA 440 | | |
| QY | 42 | KELAEDEDFGRHDKTYIRALNFYIWRKDSLNOAEANFFIEAKAA-----SSNWV 91 | | |
| Db | 441 | NQLAEIESSAQ---YKARVAL---ENDDRSEEEKYTAVORNSSEREHGSINTRENKYL 493 | | |
| QY | 92 | PKAHADPDTLAWS---KEPPRAATAG-----QQWALQTVLLVL 127 | | |
| Db | 494 | PPQQRNREVISWGSGRQNSFRMCQPSGSGMFSRSTSHTSDFNPNSGSDQ-----RVV 545 | | |
| QY | 128 | NRFPMPNNPTP---GRFTQFTLSPGSLRSPTSTNTKRDKEPANVTADPPKSLRFSAT 184 | | |

| | | | | | |
|--|-----|---|---|-----------|-----|
| Db | 546 | NGGVFWPSPCSPSPSPSPSRVQSGFNPLPPRAATPTR--- | PPSR--- | PPSP-SRPS | 596 |
| QY | 185 | GPPIHGAIP.LKFPDPVNTGSKR----- | PSLESENLOCTIKRAGKLSDNV---- | 230 | |
| Db | 597 | HPSAHGS----- | PAPVSTMPKMSSEGPPRMSPKQRPNRHRVSAGRGSISSGLEFVS | 650 | |
| QY | 231 | -----AAAAAPVPIASALDKVPTRRHANTDPATGCHRRADQVDSPTDSQTSYSGSVFS | 286 | | |
| Db | 651 | HNPPSEANTPPVARTS----- | PSGGT--WSSVVS | 677 | |
| QY | 287 | A----- | CRNQSTQSSFPAPSQPREKRPVDATVFEAGHLIESPSKGRTRTKSHINQ | 339 | |
| Db | 678 | GVPRLSPKTHPRSPRQNSINTSGFVLASPOAGIIPTEAVAMP.IPAASPTPASPSNR | 737 | | |
| QY | 340 | PLSSSSQGETS-FSTYIESFPSSGEGAIPEPGRSNGLAESASQVHVAPVVAARL | 398 | | |
| Db | 738 | AVTPSEAKDSRLQDQRQNSPAGNKENIKNET-SPSFSKAENKGIS | 790 | | |
| QY | 399 | RNI--WPFPKWLHEAPLAVAVEVTR.LFMCKVDLEDESLGLKYDPSWSTARDVTDIWK | 456 | | |
| Db | 791 | KQIDDLKFFONDPLRQSSSTESMDQ----- | LNKNREGSKSRDLI----- | 831 | |
| QY | 457 | LYRLDAFCKFPFKPPNDNVPTAMTGNFESKGSVVLSAVLDYNDPNSPTAPLVIV-- | 513 | | |
| Db | 832 | -----KDKIPSADKSIENSNCNCSGSS----- | KNPSISPSILSNT | 872 | |
| QY | 514 | KLKPLMFEQGRLT----- | RRQGRFRFFILIPS | 542 | |
| Db | 873 | HKRGSEVTSQGVQVSSSPACKQKDKBEKDAABQVRKSTLNPNAKEFNRSFSQ---- | PK | 929 | |
| QY | 543 | PTSTSPSPVPPVVSQPGAVEBVIQWLTMGQHSVLGRQWRAFFAKDAGYRKPLREQLRAE | 602 | | |
| Db | 930 | PSITFTSPRQAQPSP----- | SMVGHQ----- | 951 | |
| QY | 603 | DKPKIKERVHFFAETGITFRPDVKTRSVVPAEEVE-- | QRTFKVSOQLDWLLOLDNN | 660 | |
| Db | 952 | QTPYTPQV----- | CFAPNMYPVPVSPGVQFLPYIPMPVPMVQ----- | 992 | |
| QY | 661 | TWQPHLKFSLRIQLGLSKTYAINTLEP--HQIRHKTDLISPS | 701 | | |
| Db | 993 | -----AKTYRAVENPMQQRQDQHQSAMHPA | 1019 | | |
| RESULT 9 | | | | | |
| US-09-707-919A-19 | | | | | |
| ; Sequence 19, Application US/09707919A | | | | | |
| ; Patent No. 6623927 | | | | | |
| ; GENERAL INFORMATION: | | | | | |
| ; APPLICANT: Council of Scientific and Industrial Research | | | | | |
| ; TITLE OF INVENTION: Method for detection of human spinocerebellar ataxia 2 | | | | | |
| ; TITLE OF INVENTION: Gene variants | | | | | |
| ; FILE REFERENCE: US 443 | | | | | |
| ; CURRENT APPLICATION NUMBER: US/09/707,919A | | | | | |
| ; CURRENT FILING DATE: 2003-02-10 | | | | | |
| ; NUMBER OF SEQ ID NOS: 20 | | | | | |
| ; SOFTWARE: PatentIn Ver. 2.1 | | | | | |
| ; SEQ ID NO 19 | | | | | |
| ; LENGTH: 1312 | | | | | |
| ; TYPE: PRT | | | | | |
| ; ORGANISM: Homo sapiens | | | | | |
| US-09-707-919A-19 | | | | | |
| Query Match 1.9%; Score 139.5; DB 4; Length 1312; | | | | | |
| Best Local Similarity 17.6%; Pred. No. 0.02; | | | | | |
| Matches 145; Conservative 98; Mismatches 265; Indels 315; Gaps 34; | | | | | |
| QY | 10 | NSPVEIINRLNDY----- | NLGCQVADTTLTPH----- | RR | 41 |
| Db | 361 | NEELEALENDVSGNDPMDPMFYNEENGVVSYDLSGSIYVPLERDSEBFLKREARA | 440 | | |
| QY | 42 | KELASDEDFGRHDKIYRALNFLYWRKDDSLNOAEANFTIEAKAA----- | SSNWV | 91 | |

```

441  NQLAEETSSAQ-----YKARVAL-----ENDDRSEEEKYTAQVNSREBCHSINTRENKVI 499
QY 92  PKAHADPTLPWS---KEPPRAATAG-----QWALQTVLLEVL 127
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 494  PFGORNREVISWGSRQNSPRMGQPGSGMPSRSTSHTDNPNSGSDO-----RV 545
QY 128  NREFWPPNNPTP--GRTFGTGLSPGSLSRPTSTNTKRKDEPANVTAFDPPKRSLTRSAT 184
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 546  NGGVFWSPSCDPSRPPSRYPQSGPNSLPRAATPTP---PFSR-----PPSRP-SRPPS 596
QY 185  OPPTHGAAILPKFPDPVNTGSKR-----PSLESENLAOCTKRAKGLKLDNV--- 230
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 597  HPSAHGS-----PAPVSTPMKMSSEGPPRSPKQHRPHRNVSAGRGSISSGLEFVS 650
QY 231  ---AAAAAAPVPIASALDKVPTRRHANTRDPTATCHRRADQVDSFDSQGSYSSVFS 286
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 651  HNPSEARATPPVARTS-----PSGGT--WSSVVS 677
QY 287  A-----CRNQSTQSSFEAPPSPQFREKRPVDAVTFBAGHLIBSPSKGRTKTKSHIDNQ 339
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 678  GVERLSPKTHPRPRQNSIGNTPSGFVLASPOAGIIPTEAVAMVITPAASPTPASPASNR 737
QY 340  PLSSSSQOGETS-FSTYYESPSGGGCGATPEPSRNSGLARSEASARQVQVHAPVVAARL 398
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 738  AVTSSSAKOSRLQOQNSPAQNKENIKPNET-SPSFSKAENKGIS-----PVSEHR 790
QY 399  RNI--WPKFPKWLHEAPLAWAVEVTRLFMHCKVDLEDESLGKYDPSWSTARDVTDIWK 456
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 791  KQIDDLKFKNDPRLOPSSSTSEMSDQ-----LNKREGEKSRDLI----- 831
QY 457  LYLRLDARGKPPPKPNDYFVTAMTGNFESKGSAVVLSAVLDYNDPNSPTAPLV--- 513
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 832  -----KDKTPESAKDSFIENSSNCTSGS-----XENSPSISPSILSWTE 872
QY 514  -KLKPLMFEQOQRLT-----RREGRDFFPHILIPS 542
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 873  HKRGPEVTSQGVQSSPACQKQKDXEKKDAAEQVRKSTLNPNAKEFNPRSFQ---PK 929
QY 543  PTSTSPSPVPVVKQPGAAVEVIQWLTMGHSLVGHQWFAFPKADQAGYKPRERQLRAE 602
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 930  PSTTPTSPRQOQSPS-----SNVGHQ----- 951
QY 603  DPKPIIKERHVFHAETGITERPDPVKTSTRVFAEEPVE--QRTEFKVQMLDWLLQLDNN 660
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 952  QPTPVYTPV-----CFAPNMVYFVPVPGVQPLYPMTMPVQ----- 992
QY 661  TWQPHLKLFSRIQLGLSKTYAINTLEP--HQRHHKTDLLSPS 701
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 993  -----AKTYAVNPVPOQRQOQHQSAMMHPA 1019

```

REPLY 10
US-09-083-268-3
Sequence 3, Application US/09083268
Patent No. 6673535
GENERAL INFORMATION:
APPLICANT: Pulsz, Stefan M
TITLE OF INVENTION: NUCLEIC ACID ENCODING SPINOCEREBELLAR
SEQUENCE 3, APPLICATION: ATAXIA-2 AND PRODUCTS RELATED THERETO
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Muestring, Raasch & Gebhardt, P.A.
STREET: 119 No. 6673535th Fourth Street
City: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/083,268
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/727,084
FILING DATE: 08-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: McCormack, Myra H
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 232.00010101
TELEPHONE: 612/305-1220
TELEFAX: 612/305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-083-268-3

Query Match 1.9%; Score 139.5; DB 4; Length 1312;
Best Local Similarity 17.6%; Pred. No. 0.02;
Matches 145; Conservative 98; Mismatches 265; Indels 315; Gaps 34;

QY 10 NSPVEIIRNLNDY-----NLGQVADTTLTPH-----RR 41
DB 381 NEELEALDVSNGMDPDMFRYNEENYGVSTYDSSLSSTVPLERDNEEFLEKREARA 440
QY 42 KELEAEDDFGRHDKIYRALNFIYRKDLSLQACANFPIEAKA-----SSNWV 91
DB 441 NQLAEIEISQA-----YKARVAL-----ENDDRSEEEKYTAQRNRSEREGHSINTENKVI 493
QY 92 PKHADPDLPLWS-----KEPPRAATAG-----QWALQTVLLVL 127
DB 494 PQQRNREVLSGSGRQNSFRMGQPGSGMSRSTSHTSDFNPNSGSDQ-----RWV 545
QY 128 NRMPDPNNTP---GRTFGTSGSGLSRPTSTNTRKDEANVTADPKKSLTRSAT 184
DB 546 NGGVWPSPCPSPSPSPRSGGNSLPPRAATPTR-----PSPR-----PSPRP-SRPPS 596
QY 185 GPPIHGAALPLKPPDPVNTGSKR-----PSLESENLCQTKRAKGLSDNV-----230
DB 597 HPSAHS-----FAPVSTMPKMSSEGPPRSPKARHNRHVSAGRGSISSGLERVS 650
QY 231 -----AAAAAPPVIAALDKVPTRHANTDPTATGHRADQVDSFDTSGTSGSYVS 286
DB 651 HNPPSEATPPVARTS-----PSGGT---WSSVVS 677
QY 287 A-----CRHNSQTTQSSFEAPPSPQREKRPVDATVFEAGHLIESPGKGRITKSHDNQ 339
DB 678 GVERLSPKTHRPSPRQNSIGNITPSGVLASQAGIIPTEAVAMPIPAASPTPASPSNR 737
QY 340 PLSSSSQGETS-PSTYIEFPSSGGEGAIPEPSRNGSLARSEESARSQVQVHAPVVAARL 398
DB 738 AVTFSBEAKDSRLQDQRQNSPAGNKENIKPNET-SPSFSKAENKGIS-----PVVSEHR 790
QY 399 RNT--WPKPCKLHEAPLAVAVEVTRLFMHCKVDLEDESLGKYDPSWSMTARDVTDIWKI 456
DB 791 KQIDDLKFKVNDRLQPSSTSEMDQ-----LNKREGEKSRDLI-----831
QY 457 LYRLDAFRGKPFPEKPPNDVFTVMTGNFESKGSVAVLSAVLDYNDPNSPTAPLYLV---513
DB 832 -----KDKIEPSAKDSFIENSSNCTSGSS-----KPNSPSISPSILSNT 872
QY 514 -KLKPLMFEQGCRLT-----RRFGPDFFELIPS 542
DB 873 HKKGPEVTSGQVOTSSPACQKDKKEKDAABQVRKSTLNPNAKEFNPSFSQ---PK 929
QY 543 PTSTSPVPFWVSKQPGAVEEVIQWLTMGQHSVLGRQWRAFFAKDAGYKPLREFQLRAE 602
DB 930 PSTTTPSPRQAQPSF-----SMVGHQ-----951

QY 603 DPKPIIKERVHFAETGTFRDPDVFKTRSVVPAEPEVE--QRTFKVQSOMDLWLQDNN 660
DB 952 QPTPVYTPV-----CFAPNMYTPVPSGVQPLYPFIMTPMPVQ-----992
QY 661 TWQPHLKLFSRIQLGSLKYAINTLEP--HQIRHHXTDLSPS 701
DB 993 -----AKTYRAVENMPQQRQDGHQSAWHPA 1019

RESULT 11
US-08-976-255-11
Sequence 11, Application US/08976255
Patent No. 6136581
GENERAL INFORMATION:
APPLICANT: Jono, Keith E.
APPLICANT: Plowman, Gregory
TITLE OF INVENTION: KINASE GENES AND USES
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,255
FILING DATE: No. 6136581ember 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,675
FILING DATE: No. 6136581ember 22, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 229/182
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1384 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-976-255-11

Query Match 1.9%; Score 139.5; DB 3; Length 1384;
Best Local Similarity 21.1%; Pred. No. 0.023;
Matches 215; Conservative 114; Mismatches 371; Indels 319; Gaps 57;

QY 13 VESIIIRNLNDYNLGQCVAD---TTLTPHRRKELAESDEDFGRHDKIYRALNFIYRKD 69
DB 478 VTSTSGNLNFEYKWEAGRGAEPATLSPGRTARLQELCAPDGPVGVVPLS-----AHS 533
QY 70 DSNLQAEANFFIEAKAASNNWPKAHADPTLTPWSKEPPRAATAGQOWALQTVLLEVLNR 129
DB 534 PSLG---SEYFIRLEEA---PAAGHDPDCACAPSPP--ATADQD-----571
QY 130 FMPPNNTPORTGRTILSGP-SGLSPTSTNTKDE-PAIVTFADP--PKRSLTRSATG 185
DB 572 -----DSDSGSTAASLAMEPFLGHGPPVDVPMWRGDPYPRRSIARDPLCPSPSPSPA-G 625
QY 186 P--PIHGAA-----IPLKFPDPVNTGSKRPSLESENLCQTKRAKGLSDNVAAA 233

Db 626 PLSLAEGGADADGVAARCPAFFEDPLGTSP-----LSS 661
Qy 234 AAPVPIA--SALDKVPTERRHANTDPTATGHRRAQDVSDTSOGTSYGS--VFAC 288
Db 662 GAPPFLTGEDELEEVGARAAQ-----RGHRSNV--SANNNGSRCPESWDVVSAGC 713
Qy 289 RHNSQTQSFEAPPQPREKRPV-----ATVFEAG-----HL-----IE 324
Db 714 --HAECSPKPKOTPRASPPGFGEPPLGLQAASQEGCCGFLPILCSAQGLAPCLV 771
Qy 325 SPKSGRTTKS--HINDQP--LSSSOGTSFSTYVESFPSSGGEGAIPEPSSNGLARSE 380
Db 772 TSWTETASSGGDHPQAEKPLATEAGTTGPRPLPSPVPSQEGA--PLPS-----EE 823
Qy 381 ESARSQOVHAPVVAARLANIPKPKMLHEAPL-AVAMEVTRLMHCKV-----429
Db 824 ASA-----PDADALPDGETPATGGEVSAIKLASALNGSSSPEVE 864
Qy 430 -----DLEDESLGUKYDPS-----WSTARDVTDIWKTLYRL-----DAFRGKFPPEKPPN 474
Db 865 APSEDEDTAEATSGIFTDTSSDGLQARRPDVVPAFRSLQKQVTPDLSLSDIPSSASD 924
Qy 475 ---DVFVTWMTGNFESGSAVLSAVLDYNPNSTAPLYLYKLPLMFEOGCRLTRFG 531
Db 925 GGYEVSPSATG--PSGGQPRALDS--GYDTENY-ESPEFVLK---EAQSGCE-----969
Qy 532 PDREFFILI---PSP-TSTSPSPVVSQKQCAVEEVIQWLTMGHSLVGRQWRAFFAK 586
Db 970 POAPAEASEGEGPGPETRLSTSLSGLNEKNFVRDSAYFSDLEAEATSGPE-----KK 1024
Qy 587 DAGYRPLAEFQIARPDPIIKERVHFAETG-----ITFRPDVF-KTRSVVPAEEVPE 640
Db 1025 CGDRAPGPELGL-----PSTGQPSQVCLRPGVSGEAGQSGGPE---1064
Qy 641 QRTFEPKVSQWMLDQLDNNWTWPHLKLFESRIQLGL-----SKTY 680
Db 1065 -----VLPLLLQEGSSEP-----STCPGLVPEPPPOGPAKVRPSPSCSQF 1110
Qy 681 AIMTLEPHQIRHKTLLSPSG--TGEVNDGVGRMSRVAKRIVDLGLGVPAVOGR 738
Db 1111 FLTPVPLRSEGNSSSEFOGPGLLSGFAPQK---RMGFGTTPRALPLALPGLPAALGR 1167
Qy 739 FGSAGMWDVDDTDGDEWIETYPQKWECDV-----DXHQTLEVRSVASLKSAG 793
Db 1168 -----PEEESESESEDEBLRCYSQEPSDESEEAAPVAVVVAESQA- 1214
Qy 794 LNLQLLVLEDRARDKVKRQRAIGDLINDLQRFQSEKHALNRPVFRQWVYESYSSRA 853
Db 1215 RNL-----RSLKMPSSLSETFCEDLER---KKAVSFDDVTVLFDQES--P 1258
Qy 854 TRVSHGRVPLAGLPDSQETLNLNMGDFDPKQKYLQD-----LAWD 897
Db 1259 TR-ELGE-PP-----PGAKESPTFLRGSPGSPAPNRPQADGSPNGSTAEBGGGFAMD 1311

RESULT 12

US-09-845-917A-3
; Sequence 3, Application US/09845917A
; Patent No. 6653529
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3

LENGTH: 1528

; TYPE: PR1
; ORGANISM: Caenorhabditis elegans
US-09-845-917A-3

Query Match 1.9%, Score 139, DB 4; Length 1528;
Best Local Similarity 21.3%, Pred. No. 0.03;
Matches 156; Conservative 81; Mismatches 248; Indels 246; Gaps 33;

Qy 1 MNPITPRKRNSP-----VEEIIINLNN-----DY--NLGLQCVADITLTTHRRKELAESD 48
Db 46 INVIVPINEFPAPTKKLAKITSLNGLGLETCLDYLNKLGDC-----SKLTKD 94
Qy 49 EDFGRHDKIYRALNFLYRKDDSLNQAEANFFIEAKAASNNWVFKAHADPTLPWSKEP- 107
Db 95 IDSNGLAVLQLL-FLISTYVKQLRQ-----LKKDQKLEQLPTSIMPP 137
Qy 108 -----PRATAGQWALQTVLLEVNRMFPNPNTPGRTFGTSLGPGSLRPTSTNT 160
Db 138 AVSKLPSPRVATSAATN-----PNSNFPQMSSTRLQTPQSRISKIDSKI 185
Qy 161 KRDEPANVTADPPKSLTRSATGPIHGAAIPLKFPDPVNTGSKRP-----208
Db 186 GIKPKTSGL--KPPSSSTISSN-----NTNSFPSSRSSSGNNVGS 224
Qy 209 -----SLSS-----ENLNQCT-----KKAOKGL-----SDNVAAAAPPV 238
Db 225 TISTSAKLSSESTYSISNLNRPTSQLKPSQPOTQLVRVATTTKIGSSKLAAPKAVST 284
Qy 239 PIASALDKVTRHANTRDPTATGHRRAQDVSDTSOGTSYGSVFSACRHNQSTQSS 298
Db 285 PKLASVTIAGKEPDNSGGGGGMLKLFLSSKNPSSSSNSPQTRKAAAVPQQOTLSK 344
Qy 299 FEAPPS---QPREKRPVDATVFEAGHLIESPSKGRITK-----SHINQPLUSS 344
Db 345 IAAFPVSGLPKPTSKLSAT-----SMKLCPTKVSRYKTDAPISQODSKRCSKS 395
Qy 345 SQGETSFTYVESPP--SSGEGALPEPSRNGLARSEESARS--QVOVHAPVVAARLNI 401
Db 396 SEESGAGVAGNSPTSSSTEGSLSMHSTSKSTDEKSPSDDTLNLSIYTRQPI 455
Qy 402 --WPKFP---KWLHEAP--LAVAMEVTRLFPMHCKVDLESLGLKYDPSWSTARDVDIW 454
Db 456 AATFVSPNIIINKPVEEKFTLAVK-----GVK-----STAKK-----486
Qy 455 KTLRLDAFRGKFPPEKPPNDVFT-----AMTGN--FESKGSAVVLSAVLDYN 501
Db 487 -----DPPFAVPPRDTQFTIGVSPINAHKKLTNDPVISKEPEKLCQMSIDT 535
Qy 502 PDNSPTAPL-YLVKLPKLPMEQGRCTRREFGPRFFELIPSTSTSPSPVVPVSKQPGA 560
Db 536 TDVPPPLPLKSVVPLKWSIRQ-----PPTVDLLKQKGTISP-VKSGFYEQSSA 584
Qy 561 VEEVI-----QWLTMGHSLVGRQWRAFFAKDAGY-----RKPLEEQ 598
Db 585 SEDSIVAHASAQVTPPTKTSNHSLERRMGKNKTSBSSGYTSDAGVAMCAKMEKLEYD 644
Qy 599 ---LRAEDPKP 606
Db 645 DMTRAQNGYP 655

RESULT 13

US-09-845-917A-4
; Sequence 4, Application US/09845917A
; Patent No. 6653529
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR


```
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1583
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-845-917A-4

Query Match
Best Local Similarity 1.9%; Score 139; DB 4; Length 1583;
Matches 156; Conservative 81; Mismatches 248; Indels 246; Gaps 33;

Qy 1 MNPTTRKNSP-----VBEINRLN-----DY--NLGQVADTTLTPHRRKELASD 48
Db 46 INVIVINEFSPAFTRKALAKITSNLDGLTCLDYKLNGLDC-----SKLTSTD 94
Qy 49 EDFGRHDKIYRALNFIYRKDDSLNQAENAFTEAKAASSNVVPKAHADPTLPMSKEP- 107
Db 95 IDSGNLGAVIQLL--FLLSTYKQLRQ-----LKKDQKKLEQLPTSINPP 137
Qy 108 -----PRAATAGQWALQTVLLEVINRFPMPNPTPGTGRILSGSGLSRTSINT 160
Db 138 AVSKLPSRPRVATASATN-----PNSNFPQMSTSRLOTPOSRIKIDSSKI 185
Qy 161 KRDEPANTVFADPPKRSLTRSATGPIIHGAALPLKFPDPVNTGSKRP----- 208
Db 186 GIPKTSGL---KPPSSSTSSN-----NTNSFRPSRSRSGNNVGS 224
Qy 209 -----SUBS-----ENLNQCT-----KRAKGL-----SDNVAAAAPV 238
Db 225 TISTSAKLSBSTSYSSINLRPTSQLQKPSRPQTLVRVATTTKIGSKLAAPKAVST 284
Qy 239 PIASALDKVTRRHANTROPTATGHRRAQVDSFDSQTSYGSSVFSACRHNQSTQSS 298
Db 285 PKLASVTIOAKGPNSSGCGGGMKLKLFSSKNFSSSSNPQPRKAAAVPQOOTLSK 344
Qy 299 FEAPPS---QPREKRPVDATVFBAGHLIESPSKGRITK-----SHIDNQLSSS 344
Db 345 IAAPVKSGLKPPSTKLSGSA-----SMSKLCIPKVSRYKTDAPITISQDSKRCSSK 395
Qy 345 SQGTSSTSYESFP--SSGEGAIPEPSRNGIARSESARS--QVOVHAPVVAALRNI 401
Db 396 SEESGAGPNSTSPSTSSSTEGSLMHSTSSKSTSDKSPSSDDUTLNASIVTAIRQI 455
Qy 402 --WPKFP---KWLHEAP-LAVAMEVTRLFMHCKVDLEDESLGLKYDPPSSTARDVTDI 454
Db 456 AATPVSENIINKPVEEKPTLAVK-----GVK-----STAKK----- 486
Qy 455 KTLRLDAPGKPEPEKPPNDVFT-----AMTGN--FESKGSAAVLSAVLDYN 501
Db 487 -----DPPPAVPPRDTQTTIGVSPIMAHKKLNDPVISEKPEKLGNSIDT 535
Qy 502 PDNSPTAPL-VLVKLKPLMEQCGRLTRRPGDRFPFEILIPSTSPSPVPFVSKQQA 560
Db 536 TDVPPPLPSKVVPLKMTSIRQ-----PPTVDLLKQKITSP-VKSGFYEQSSA 584
Qy 561 VEEVI-----QWLWQGSLSVGRQWRAPFAKDAGY-----RKLPRBQ 598
Db 585 SEDSIVAHASAQVTPPTKTSNHSLERRMGKNKTSSESGYTS DAGVAMCAKVRKLEK 644
Qy 599 --LRAEDPKP 606
Db 645 DMTRRAQNGYP 655

; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11704
; LENGTH: 1768
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11704

Query Match
Best Local Similarity 1.9%; Score 137; DB 4; Length 1768;
Matches 196; Conservative 142; Mismatches 347; Indels 326; Gaps 53;

Qy 407 KWLHEAPLAVAMEVTRLFMHCK-----VDLEDESLGL--KYD--PWSSTARDV 450
Db 778 KSLFLKPYGEGEALPHFIYQERDDLSSPVRTQVIELIDETWGLINIHDELPRASEKVG 837
Qy 451 TDWTKTILRLDARGRKPPPEKPPNDVP--VTANTGNFESKGSAAVLSAVLDYNDPSPTA 508
Db 838 LSLWLEKVKDXYRGEGQRKKILNALLKVSTA.KDDP-----DELMKQ 880
Qy 509 PLVIVLKLPLMFEQCGRLTRRFGDRFFFEILIPSTSTSPSPVPVSKQGAVEEVI--Q 566
Db 881 DVFTSKIKP-----RRLS-----YVDELSSALIGFNVPMCKFYDFVVKLSMHE 926
Qy 567 WLTWQGSLSVGRQWRAPFAKDAGYRKLRFQRLRADPKPIIKERVHFFRAETGITRPDV 626
Db 927 WLLSESEE-----NDYGY-----GSYRVDVSESGFLDTER-DFFPAG----- 963
Qy 627 FKTESVVPABEPVORTEFKVQSWMLQLDNNWTQPHLKL-FSRIQLGLSKT-YAINT 684
Db 964 -----AKPFFKYLQFAPRLGLDPIRLCNLTAKYSESDFSKTDNKEDTIYAFET 1015
Qy 685 LEHQIRHHKTDLLSPGTEGMNDGVGRMSRSVAKRI-----RDVLGLGDVPSAQGR 738
Db 1016 VA-----KQVDI-----TLNDG-----TVKQYASPHLWKGYRGOSTLFLYLLQSA 1055
Qy 739 FGSAGKMWVID-VDDTDGDEWIETPSQKWCEDFVDKHQRTLEVRSAVELSAGLN-- 795
Db 1056 LMALEN-WLVEYVENCCKNEID-----WIDYVLRSSNSVMPTSVLSSVATGPENKV 1107
Qy 796 ----LQLLPVLEDRARDKVMRQAIGRLINDLQROFSEOKHALN-----RPEFROW 844
Db 1108 GKAAFPPLKTDADLYHLDLIRMTQEMGG--NEM-HFFGLNRDALSKEYLEEREAALRPW 1163
Qy 845 VYESYSRATRVSHGRVFPFLAGLPD-----SOBETLNLNMSGFDPKQK 889
Db 1164 RKESLETLLRLQ-----FVNELRDILKIVDELKNEATASNEKSLRYMYHR-VDTRTWK 1217
Qy 890 YLOD-----IAWDL-----QKRCOTLKSRLNIRVGRSAIYIWIADFNG 928
Db 1218 VVEDKENDRVLFSSSELPEDLKLDQOEKNEKHAMDNVTLSNL-----WG 1263
Qy 929 VLENEVHVGFSSKFRDEE--ESFTLLSDCDVLVARSPAHPSPDIQVRVAFKPELSLK 986
Db 1264 -----KCLFDEQLLEEKYFSSYQDALIATKGL-----LSALQKEVNFPA 1303
Qy 987 DVII--FSTKGDVP.LAKKLSGGDYDGMWVCMWDPEIVDGFVNAEMPLEDLSLYLKDK 1044
Db 1304 DMAVGTTITVAACVVRDVLLELSYE-DKKW-CLEIILLESIFMHAD-----NMNGTAADHK 1356
Qy 1045 TTFKQLWASHGTGSA-----KQOTTYDMIQKSFHFALQPNFLGCMCTNYKRL 1092
Db 1357 TDY-----YGSACAFVLPKLFOLDLDSQIET-----LKFAL-----ATAL 1393
Qy 1093 CYINNSVS-----NKPAIILSSLVGNLVDQSK-----QGIYFNE 1126
```

RESULT 14

US-09-489-039A-11704

; Sequence 11704, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

Db 1394 THENLNVSAKAGVREFWMSRDAELASRCIGGIVEYARLRREDSEVRRFVHLOQAELOA 1453
Qy 1127 A--SWAQL---RRELLGG--ALSPLDPMYKSD-SWLGRGEPHTHIDYLKFSIARPAIDK 1177
Db 1454 ALEKWNDLTSPNDLIEGKFKLSYNDISLESCHSFWIH-LPMLVPICTKD-----EI 1506
Qy 1178 ELEAFNANKAAKDTEDGAHFWDPLASYYTFFKEISD-----KSRSS-----A 1221
Db 1507 QQLVNRLLMSFVDFTEYQNHRSDDDEKINHEIKKQIQCLSEHVSSRRNNFMFPKOLLV 1566
Qy 1222 LLFTTLKNRIGVEYERKGLVKNKMRDSDKDPYVRVQVYKWKCAITPEA 1272
Db 1567 LGCTKAPGGFIYSLTSLSYHVAVEKEGDYDA-----IWALWCLLAPEA 1607

RESULT 15
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivria, John
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR FILING DATE: 60/136,620
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match
Best Local Similarity 18.4%; Pred. No. 0.16;
Matches 196; Conservative 117; Mismatches 364; Indels 389; Gaps 44;

Qy 82 EAKASSNVVRAHADPDPLPMSKEPPRAATAGQWALCTVLEVLNRFMPNNTP-GR 140
Db 1125 EVSASAPGAAPTISPLHVPSSLPGPASSP-----MFIPISSPLAS 1166
Qy 141 TFGRLSGPSGLSRPTSINTK-----RKDEPANVTADPPKRSILTR 181
Db 1167 PVSSIVSVPLSSSLPISVPTTLAPASAPLTIPISAPLTVSASGALLTSVTPPLAPVP 1226
Qy 182 SATGP----- 187
Db 1227 AAPGPPSLQPSGASPSASALTGLATAPSLSSQTPGHPHLLAPTSHPVGLNSTVPAC 1286
Qy 188 ----THGAAPLPKFPDPVNTGSKRPSLESENINQCTKRAKGLSDNVAAPVPI--- 240
Db 1287 SPVLVPASALASFPFSAPNPAQAASLLAP-----ASSASQALATFLAPMAAPQTALAP 1341
Qy 241 ---ASALDKVPTRRHNRDPTATGHRADQVDSFTSGTSYGSS-----VFSACRH 290
Db 1342 SPAPPLAPLPLVLAAPGAPVLASSQTPVPMAPSFTGTSLASASVPAPTPVLA----- 1397
Qy 291 NOSTTQSSPEAPPSOPREKRPVDATVFEAGHLIESPSKGRITKSHIDNQPLS-SSSQGET 349
Db 1398 -PSSQTMLPAPVPLSPFPSPASTQILALAPAL--APTGGSSPS-----QTLSTGTPQG 1450
Qy 350 SFSTYYESPPSGGGAIPERSRNGLARSEBSAQVQVHAPVVAELRNINWPKFKWL 409
Db 1451 PFPT--QTLSLTPASSSLVTPAQTLSLAFGPPGLGPTQTLISLAP--APPLAPASVPGAPA 1506
Qy 410 HEAPLAVAVEVTRFLMHCKVDLEDSEGLKDYDPSSTARDVDIWKTLYLDAFRGKFPF 469
Db 1507 HILTAPASSSASLAPASVQT-----LT-SPA-----PVP 1537
Qy 470 EKPP-----NDVFVTAMTNGFESKGSVAVLSAVLDYNDPNSPTAPLYLVKLXPLMFEQGR 525

Search completed: March 29, 2004, 08:50:39
Job time : 33 secs

Db 1538 TLGPAAQTTLALPASTQSPASQSLVWSA-----SGAAPLFTVMVSRSLPVSKDEPDT 1591
Qy 526 LTRFPGDPDRFEILLIPSTSTSPSV-----PPVVSQOP----- 558
Db 1592 LTLRSGP-----PSPSTATSGGPRPRRQPPPPRSPFFYLDLSLEKKRQSRSELE 1643
Qy 559 -----CAV-----BEVIQMTWQO--HSLVGRQ-----WRAPFAKDAGYRKP 593
Db 1644 RIFQSEAHGALAPVYGTVEVLDFTLFPQVVASPIGRSPGSPSHPTWT---YTEAAHRAV 1700
Qy 594 LRSEFO-----LRAEDPKPIIKERVHFFAE'GITFRDPDVPK 628
Db 1701 LPQORLDQLSEIERFIVMPVPEAPPSLHACHPPWLAPOAAFOEQ-----L 1751
Qy 629 TRSVVPAEPEVQO-----RTEFKVQOMLQDNNNTWPHLKLFSRIQLGLSKTYAIM 683
Db 1752 ASELWPRARPLHRIVCNMRTQF-----PDLRLI-QYDCGKLQTLAVL 1792
Qy 684 TLEPHQIRHHKT-----DLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLG 729
Db 1793 -LRQLKAEGRHVLIPTQMTRMLDVLQOFLTYHGLYLRDGGSTRVEQR----- 1839
Qy 730 DVPSAVQGRFGSAGKMWV-----IDVDDTG-----DEDWIETTPSQKWECDVF 773
Db 1840 ---QALMERENADKRIFCFILSTRSGGVNLTGADTVVYDSDNWPTMDAQADRCHRI 1896
Qy 774 DKHQTLEVRVSASELSKAGLNQLLPVLEDRARDKVKRQALGDRLI---NDLQRPSE 830
Db 1897 GQ-TRDVHIYRLISE-----RTVEENILKANOKRMGLDVAIEGNGFTTAYFKQ 1944
Qy 831 Q--KHALNRPEPRQWVYESYSSSRATRVSHGRVPFLAGLPDSQEBTLNFMNLSGPDPKQ 888
Db 1945 QTIRELFDMELE-----EPSSSSVP---SAPEEEETV-----ASKQ 1978
Qy 889 KYLQDIANDLQKXKCDTLKSKLNIRVGRSAYIYMIADFWGLVEENE 934
Db 1979 THI-----LEQALC-RAEDEEDIRAATQAKAQVAE---LAEFNE 2014

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 02:06:52 ; Search time 8802.43 Seconds
(without alignments)
6903.427 Million cell updates/sec

Title: US-09-913-878A-2
Perfect score: 7397
Sequence: 1 MNPITPRKRNSVPEIIRL.....YEVLGDDDDFDGIGTNGNDY 1402

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082008_8366/app_query.fasta_1.2254
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878 -GCN1_7684 @runat_29032004_082008_8366 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.hcg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.hcg.hum.*
31: em.hcg.lrv.*
32: em.hcg.other.*
33: em.hcg.mus.*
34: em.hcg.pln.*
35: em.hcg.rtd.*
36: em.hcg.nam.*
37: em.hcg.vrt.*
38: em.sy.*
39: em.hgo.hum.*
40: em.hgo.mus.*
41: em.hgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|-----------|--------------------|
| 1 | 7397 | 100.0 | 4206 | 8 | NCF133528 | AJ133528 Neurospor |
| 2 | 7397 | 100.0 | 8045 | 6 | AX034425 | AX034425 Sequence |
| C 3 | 348.5 | 4.7 | 165701 | 8 | AP004357 | AP004357 Oryza sat |
| 4 | 328 | 4.4 | 3505 | 8 | NTA011576 | AJ011576 Nicotiana |
| 5 | 328 | 4.4 | 3591 | 6 | AX204840 | AX204840 Sequence |
| C 6 | 306 | 4.1 | 82584 | 8 | NCB1383 | BX284762 Neurospor |
| 7 | 299 | 4.0 | 6847 | 8 | AY029284 | AY029284 Neurospor |
| 8 | 298 | 4.0 | 5992 | 3 | DDI314910 | AJ114910 Dictyoste |
| 9 | 297.5 | 4.0 | 3731 | 6 | ARI45904 | ARI45904 Sequence |
| 10 | 297.5 | 4.0 | 3731 | 8 | LERDRP | Y10403 L.esculentu |
| 11 | 296.5 | 4.0 | 6688 | 8 | AF468822 | AF468822 Diaphorte |
| 12 | 293.5 | 4.0 | 5731 | 3 | DDI314909 | AJ314909 Dictyoste |
| 13 | 290 | 3.9 | 3807 | 8 | AY148431 | AY148431 Arabidops |
| 14 | 289.5 | 3.9 | 4013 | 8 | AF239718 | AF239718 Arabidops |
| 15 | 289.5 | 3.9 | 6863 | 6 | AX204839 | AX204839 Sequence |
| C 16 | 289.5 | 3.9 | 80367 | 8 | AC012329 | AC012329 Arabidops |
| C 17 | 289.5 | 3.9 | 104204 | 8 | ATT9C5 | AL132964 Arabidops |
| 18 | 285.5 | 3.9 | 1538 | 8 | ATH011977 | AJ011977 Arabidops |
| 19 | 284.5 | 3.8 | 5280 | 8 | AF268093 | AF268093 Arabidops |
| 20 | 281 | 3.8 | 6987 | 8 | AY049072 | AY049072 Diaphorte |
| C 21 | 277.5 | 3.8 | 132716 | 2 | AC105931 | AC105931 Magnapor |
| 22 | 273.5 | 3.7 | 3524 | 8 | AF443073 | AF443073 Phomopsis |
| 23 | 268 | 3.6 | 4658 | 8 | SPAC6B12 | Z98533 S.pombe chr |
| 24 | 265 | 3.6 | 7178 | 8 | AF500110 | AF500110 Neurospor |
| 25 | 261.5 | 3.5 | 29132 | 3 | CEF10B5 | Z48334 Caenorhabdi |
| 26 | 261 | 3.5 | 6204 | 8 | AF411016 | AF411016 Neurospor |
| 27 | 253.5 | 3.4 | 2639 | 8 | AK101440 | AK101440 Oryza sat |
| 28 | 247 | 3.3 | 132699 | 8 | AC006917 | AC006917 Genomic s |
| C 29 | 242.5 | 3.3 | 121337 | 8 | OSJN00104 | AL606653 Oryza sat |
| 30 | 236.5 | 3.2 | 4978 | 3 | AF117611 | AF117611 Dictyoste |
| 31 | 236 | 3.2 | 5025 | 3 | AF159144 | AF159144 Caenorhab |
| 32 | 230 | 3.1 | 5208 | 3 | AF159143 | AF159143 Caenorhab |
| C 33 | 225 | 3.0 | 74406 | 8 | F2P3 | AF080120 Arabidops |
| 34 | 225 | 3.0 | 108598 | 8 | ATT22B4 | AL049876 Arabidops |
| 35 | 225 | 3.0 | 198301 | 8 | ATCHRV31 | AL161531 Arabidops |
| 36 | 218.5 | 3.0 | 4300 | 3 | AF293414 | AF293414 Giardia i |
| 37 | 217.5 | 2.9 | 6103 | 3 | AF393743 | AF393743 Giardia i |
| 38 | 211.5 | 2.9 | 299925 | 1 | AP005045 | AP005045 Streptomy |
| 39 | 209.5 | 2.8 | 3628 | 8 | AK110169 | AK110169 Oryza sat |
| 40 | 199.5 | 2.7 | 4579 | 8 | PHY011979 | AJ011979 Petunia h |
| 41 | 196 | 2.6 | 308050 | 1 | SC0939124 | AL939124 Streptomy |
| C 42 | 195.5 | 2.6 | 17589 | 1 | AE004862 | AE004862 Pseudomon |
| C 43 | 195 | 2.6 | 11908 | 1 | AE005822 | AE005822 Caulobact |
| C 44 | 194.5 | 2.6 | 31521 | 3 | CEY47H10A | AL132848 Caenorhab |
| 45 | 192.5 | 2.6 | 300350 | 1 | BX294145 | BX294145 Pirellula |

ALIGNMENTS

RESULT 1

NCR133528
LOCUS NCR133528 4206 bp DNA linear PLN 10-MAY-1999
DEFINITION Neurospora crassa qde-1 gene, partial.
ACCESSION AJ133528
VERSION AJ133528.1 GI:4803726
KEYWORDS qde-1 gene; RNA-dependent RNA polymerase.
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Peizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
1
Cognoni, C. and Macino, G.
Gene silencing in neurospora crassa requires a protein homologous
to RNA-dependent RNA polymerase
Unpublished
Cognoni, C.
2 (bases 1 to 4206)
Direct Submission
Submitted (08-MAR-1999) Cognoni C., Dept. Biotecnologie Cellulari ed
Ematologia,, Universita di Roma La Sapienza, Viale Regina Elena
324, 00161 Rome, ITALY
FEATURES
Location/Qualifiers
1..4206
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
1..4206
/gene="qde-1"
1..4206
/gene="qde-1"
/function="gene silencing"
/note="putative"
/codon_start=1
/product="RNA-dependent RNA polymerase"
/protein_id="CAB42634.1"
/db_xref="GI:4803727"
/db_xref="GOA:Q9Y7G6"
/db_xref="SPTREMBL:Q9Y7G6"
/translation="WNPTPRKNSPVBEIINRLNNDVNLGQCVADTTLTPHRRKEL
AESDDFGRHDKIYRALNFIYWRKDDSLNQAENFFIEAKAASNNWPKAHADPTILP
WSKEPRANTAQQWALQTLVLENLNRPMPNNPTGFTFGRTLSGSGLSRPTSTNT
KRKDPANVTADPPKSLTSATGPPIHGAALPLKPPDPVNTGSKRPSLESENLOC
TKRAGKLSDNVAAAAPVPIASALDKVPTRRHANTDPTATGHRADVDSDPTSQ
GTSYSSVPSACRHQNSTQSGFAPPSQPREKRPVDATFEAGHLIESPSKGTTS
HIDNPLSSSQGETSFSTYSFSSGGEAIPERSNSGLARSERASQVQVHAP
VVAARLNIPKFPKWLHEAPLAVAVEVTRFLMHCKVLEDESLGLKYDPSWSTARDV
TDIWKTLRLDAFRGKPPFEKPPNDVFTVMTGNFESKGSVAVLSVLDYNDPSFTA
PLYVLKLPFMEQCRUTRREFGRFPEILIPSTSTSPSPVPVSKPGAVEVQ
WLTMGHSLVGMQWAFPAKADAGYKPLREPOLRAEDPKPIKERVHFFAETGIFRP
DYFKTSVVPAEPEVEQTEFKVSOMDLWLQDNNWQPHLFLSRTOLGLSKTYAI
MTLEHQTHHHKTDLLSFGCTGEVMDGVGMSRSVAKIRDLGLGDPVSAVOGRG
SAKGMWIDVDITGDEMIETIYPSRKNECDFVDKHQRTLSVRSVASELKSAGNLQL
LPVEDRARKVMRQATGDRLLNDLQKQFSEQRKALNRPFVROWVYESYSSAIRV
SGRVPFFLAGLPSDEETLNLNMSGDFPKKQYLQDIAWDLQKCKDILSKLNIRV
GRSAVTYMIADFWGLNEVHVGFSSKFRDEEFTLLSDCDVLVARSAPAFPSDIQ
RVRAVFKPHELKSDVITIPSTKGDVPLAKLSGGDYGDMAMVWDPEIVDGFVNAEM
PLEPDLRYLKKDKTTFKQLMASHGTGSAKQOTTYDMIQKSFHALQPNFLGNTY
KERLCYVNNVSKKPAIILSSLVGNLVQSQKQIVNEASWAQRLRLGLGALSIPDP
MYKSDSWLGRGEPHILIDYKESIAIPALDKLEAFHNAKAKDTEGAFHMPDLA
SYTFPKEISKRSALFTTKRIGEKVEKEIGRLVKNKEMRSDSKDPYFVRNQVY
EKWATFIPAMKDSGANDYSKVRILLELSFADREMTWALLRSTAFKLYIHKSPX
VWQMGROLAYIAQWTSRPGEGALMTAFMYAGLMPDKFKTKYVARLEGDSBP
DPEYVEVLGDDDDFGIGFTNGNDY"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4206
Score: 7397.00 Matches: 1402
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-09-913-878A-2 (1-1402) x NCR133528 (1-4206)
Qy 1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleLeuArgLeu 20
Db 1 ATGAACCTTATTACTCTAGGAAGAGGATAGCCCGTCGAGAAATATATAACCGGCTC 60
Qy 21 AsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThrThrLeuThrProHisArg 40
Db 61 AATAACGACTTACAACCTGGGCTTCCAGTGTGTCGAGACACAACTCTACCCCCACCGC 120
Qy 41 ArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAla 60
Db 121 CGGAAGGAGCTGGCCGAGAGTGCAGAGATTTCGTGCGCATGACAAGATCTTACAGAGC 180
Qy 61 LeuAsnPheLeuTyrTrpArgLysAspSerLeuAsnGlnAlaGluAlaAsnPhePhe 80
Db 181 CTGAACCTTCTTACTACGCGGAGAGATGACTCTCTGACACGAGACGCAACTTCTTC 240
Qy 81 IleGluAlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThr 100
Db 241 ATCGAGGCGCAAGCTCGAGCTCGAACTGGGTGCGCCAAAGCCGACGCCCTGACAGC 300
Qy 101 LeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnTrpAlaLeuGln 120
Db 301 CTTCCGTGTGTCGAAGGAACCTCCCGCGCGGTACTGCGCGCCCAACATGGGCATTGCGAG 360
Qy 121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
Db 361 ACTGTGTGTCGAGGTGCTTAATAGTTTATGTCACCTCCCAATAACACACAGGTGCA 420
Qy 141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
Db 421 ACCTTTGGCAGAACTCTAAGCGCGCCCAAGTGCGCTGAGCGCCCAACCTCTACCAACAC 480
Qy 161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
Db 481 AAACGCAAGGATGAGCGCGCAATGTCACTTCGTGTGATCCGCCCAACCGTCGTTGACT 540
Qy 181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
Db 541 CGCTCTGCCACAGGTCTCTATTACGCGCGCGCGATACCCCTCAAGTTCCCCGATCCA 600
Qy 201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
Db 601 GTGAATACCGTTCCAAACGACCATCTCTCGAGAGTGAGAAATCTCAATCAGTGCACCAAG 660
Qy 221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValProIle 240
Db 661 CGGCGCAAGGCGCAAGCTGTCTGATAATGTTCGCGTCCCGCGCGCCCGCGCGCTATT 720
Qy 241 AlasrAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db 721 GCAGGCGCTTTGGACAGAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCACGGCG 780
Qy 261 ThrGlyHisArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db 781 ACAGGTCAATAGACGAGCGGACCAAGTGGATTCCTTTGATACATCTCAAGGCACCTTCCTAT 840
Qy 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db 841 GGTTCAGTGCTTTCAGCGCTTCCCTGACATCATCAGAGCACTACCCAGAGTAGTTTGTAG 900
Qy 301 AlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGly 320
Db 901 GCTCTCTCTTTCACAGCCCGAGAGAGCGGCTGTGATGCCACGCTCTTTGAGGCTGGA 960
Qy 321 HisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsnGlnPro 340
Db 961 CACTTGATTGAGTCTCTTAGCAAGGAAGAACCAAGTCCCAATAGTATACCGAGCCC 1020
Qy 341 LeuSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
Db 1021 CTTTCATCGTCTTCCCAGGCGTGAATCTTCGTCAGCACCTTACTATGAGTCGTTTCCAAGT 1080

QY 361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
 Db 1081 TCCGCGCGAGGGCGCAATCCGAGCGGAGTCGCTCAATAGGACTGGCTCGAGCGAA 1140
 QY 381 GluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsn 400
 Db 1141 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCGGGTGGCAGCTCGGCTCAGAAAT 1200
 QY 401 IleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluVal 420
 Db 1201 ATTTCGCGCAATTTCCCAATGGGTACACGAAGCTCCTCTCGCTGTTCATGGGAAGTT 1260
 QY 421 ThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyr 440
 Db 1261 ACCAGACTCTTATCATCTGCMAAGTAGACTTGGNAGACGAGCCTGGGCTTAAAGTAC 1320
 QY 441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeu 460
 Db 1321 GACCCCTTCTGCTACCGCGCGGATGTCACAGATATCTGGAAGACTCTCTACCGGCTT 1380
 QY 461 AspAlaPheArgGlyValProPheProGluLysProProAsnAspValPheValThrAla 480
 Db 1381 GATGCTTTCCGTGGTAAACCTTTCCAGAAAGCCGCCAACGACGCTGCTCGTACCGCA 1440
 QY 481 MetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyr 500
 Db 1441 ATGACGGGCAACTTTGAGAGCAAGGTAGTGCCTGTGTCTCTGCTGTTCTTAGACTAC 1500
 QY 501 AsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPhe 520
 Db 1501 AATCCGGCAACTCGCCCTACTCGGCCCTTTTACCTGTGGAAGCTGAAGCCGCTCATGTT 1560
 QY 521 GluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIleLeuIle 540
 Db 1561 GAGCAGGGCTGTCGACTCACCGTCTCGTTCGTCTGATAGTCTTTTCGAGATCCTTATA 1620
 QY 541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560
 Db 1621 CGCTCGCTTACAGACACCGCCAAAGTGTACCGCGGGTGGTCAGCAACCAACAGGTGCG 1680
 QY 561 ValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrp 580
 Db 1681 GTCCAGAAAGTCAATCAGTGGCTCAGATGGGCGCAACATCTCTTGTAGGCGCCCAATGG 1740
 QY 581 ArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArg 600
 Db 1741 CGCGCTTTCTTCGCCAAAGATCCCGGATACAGAAACCTCTCAGGGAGTTCAGCTCCGC 1800
 QY 601 AlaGluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIle 620
 Db 1801 GCCGAGGCCCGAAACCCATCATCAGGAGAGAGTCCACTCTTTTGGCCGACCGGCATT 1860
 QY 621 ThrPheArgProAspValPheLysThrArgSerValProAlaGluGluProValGlu 640
 Db 1861 ACGTTCGACCTGATGTGTTCAAGACGAGATCTGTCTGTCGGCAGAGGAACCTGTAGAG 1920
 QY 641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuGlnLeuAspAsnAsn 660
 Db 1921 CAACGACCGAGTTCAAAGTTAGTCAATGCTGGACTGGCTCTCTGCAACTCGAACCAAC 1980
 QY 661 ThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyr 680
 Db 1981 ACTTGGCAGCCGACCTCAAGTTGTTCCCGTATCCAGCTCGGTCTCAGTAAGACATAT 2040
 QY 681 AlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerPro 700
 Db 2041 GCCATTATGACATTGAGCGCTCACCAGATCAGACACCAAGACCGATCTTCTTTACCT 2100
 QY 701 SerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLys 720
 Db 2101 TCAGGCACCTGGCGAAGTATGATGACGGTGTAGCCCGCATGTCGCGAAGCGTGGCCAG 2160

QY 721 ArgIleArgAspValLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
 Db 2161 AGATACGGCGATGTTCTCGGTTTGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTGGT 2220
 QY 741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
 Db 2221 TCGGCCAAGGGAATGTGGGTATTCGACGTTCAGCACACAGGCGATGAGGATTGGATCGAG 2280
 QY 761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
 Db 2281 ACATACCCGTCGCGCGCAAGTGGGATCGACTTCGTTGATAAACATCAACGTACCCCTC 2340
 QY 781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuPro 800
 Db 2341 GAAGTCGGGAGCGTCTCTGAACCTGAAGTCAGCTGGTCTCAACATCAGCTGTACCT 2400
 QY 801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
 Db 2401 GTCTCTGGAAGTAGACCCAGGACAGGTGAAGATCGCCGCGCAATCGTGACCGTCTT 2460
 QY 821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
 Db 2461 ATCAACGATTTGCCACGACAGTTCAGCGAGCAAAAGCATGCTTGAATCGCCAGTGGAA 2520
 QY 841 PheArgGlnTrpValTyrGluSerTyrSerArgAlaThrArgValSerHisGlyArg 860
 Db 2521 TTTTCGCAATGGGTTTTACGAGATTTATCCAGTCGCGCAACTCGAGTCAGCCAGCCGT 2580
 QY 861 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 880
 Db 2581 GTGCCCTTTCTGCTGGGTACTCTGACAGTCAAGAGGAGACACTGAATCTTTGATGAAC 2640
 QY 881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
 Db 2641 AGTGGGTTCGATCCCAAGAGCAAAAGTACTTTCAGAGACATCGCCTGGGATCTTCAAAAG 2700
 QY 901 ArgLysCysAspPheLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTrpIle 920
 Db 2701 CGGAATGTGACACGTTGAAGTCAAGCTGAATCCGTGTGGTGCATCAGATACATT 2760
 QY 921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
 Db 2761 TACATGATTCGGATTTCTGGGTGTGCTTGAAGAAAATGAGGTTTCATGTCGGATTTCTCC 2820
 QY 941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
 Db 2821 TCAAGTTTCAGACGAGGAGGAGTCTTTTCACTCTATCGGACTGTGATGTCTCTCGTG 2880
 QY 961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
 Db 2881 GCGGATCCCGACCCCATTTCCCTAGTATATCCAAACGGGTTCGAGCAGTCTTCAAGCCA 2940
 QY 981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
 Db 2941 GAGCTCCAGCTCTCAAGGATGTAATCTCTTCTTACTTAAGGAGATGTACCGCTGTCT 3000
 QY 1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
 Db 3001 AAGAGCTATCTGGTGGAGACTACGACGCGGATATGGCTGGGTCTGCTGGGATCCGAG 3060
 QY 1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
 Db 3061 ATCGTCGATGGTTTCGTCAATGCGGAATGCCCTCTGGAGCCGACCTGTCTAGGTACCTA 3120
 QY 1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
 Db 3121 AAGAGGACAAACAGACTTTCANACAACTTATGGCTCACACGCGACGGCTCAGCGGCC 3180
 QY 1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
 Db 3181 AAAGAGCAGACTACATACGATAATGATCCAGAGAGGTTCATTTCCCTTCGCGCCCAAC 3240
 QY 1081 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 1100

Db 3241 TTCTTGGGATGTCACAAAGGCTCTGTACATCAACAATAGTGTCT 3300
 Qy 1101 AsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuValAspClnSerLysGln 1120
 Db 3301 AACAGCGCGCATATCTTAGTTCACTGGTGGAAACCTCGTCGATCAGAGCAAGCAA 3360
 Qy 1121 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuGlyGlyAla 1140
 Db 3361 GGTAATGCTTTTACGAAGCAAGCTGGGCTCAATTCGTAGGAACCTGCTTGGCGGTGCA 3420
 Qy 1141 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThr 1160
 Db 3421 TTGTCCCTTCTGACCCATGTACAGAGCGACAGTGGCTCGGCGGAGAGCGCTACC 3480
 Qy 1161 HisIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 1180
 Db 3481 CACATTATTGACTACCTGAAATTCCTCATCGCAGGCTCGGATTCAGAGCAAGCACTGGA 3540
 Qy 1181 AlaPheHisAsnAlaMetLysAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
 Db 3541 GCCTTCCCAATGCCATGAAGCGGCCAAGGATACAGAGCGGCGCTCACCTTTTGGGAT 3600
 Qy 1201 ProAspLeuAlaSerTyrTrpThrPhePheLysGluLysSerAspLysSerArgSer 1220
 Db 3601 CCGGATCTCGTCTCTACTACAGTCTCTCAAGGAGATAGGACAGATCGGATCGTCC 3660
 Qy 1221 AlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg 1240
 Db 3661 GCACCTGCTATTCAGCACTCTGAAGAACCGTATCGGCGAAGTCCGAGAGAAATATGCGCAG 3720
 Qy 1241 LeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValaAsnGln 1260
 Db 3721 TTGGTCAAAAACAGGAGATGAGACAGACAGGACCCCTACCTGTCCGCGTCAACCCAG 3780
 Qy 1261 ValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyr 1280
 Db 3781 GTTTATGMAAATGGTGGCCATCACCGCTGAGCGATGCAACAATCCGAGCAAAATAT 3840
 Qy 1281 AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
 Db 3841 GATTCTAAGGTGATCAGGTGCTGGAGCTGTCTCTCCGCGACCGTGAATGATGATACA 3900
 Qy 1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
 Db 3901 TGGGCAITGCTGAGGCTAGCACGGCTTTCAGCTGTACTACCAAGAGCCCAAGTTC 3960
 Qy 1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
 Db 3961 GTCTGGCAGATCGCGGACAGACGCTCGGTACATTAAAGCGCAGATGACGAGACGCC 4020
 Qy 1341 GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
 Db 4021 GGTGAAGCGCCCGCGGTGATGACCGCTTCATGTATCGCGCTTGATCGCGGATAAG 4080
 Qy 1361 LysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspPro 1380
 Db 4081 AAGTTTACGAAGCAGTATGTGGCGAGCTGGAGCGGATGGATCGGAGTACCTGATCCG 4140
 Qy 1381 GluValTyrGluValLeuGlyAspAspPheAspGlyIleGlyPheThrGlyAsnGly 1400
 Db 4141 CAGGTCTATGAGTGTGGCGATGATGATTTTGTATGGATTTGGTTTCACAGGAATGCG 4200
 Qy 1401 AspTyr 1402
 Db 4201 GATTAT 4206

RESULT 2
 LOCUS AX034425 8045 bp DNA linear PAT 22-SEP-2000
 DEFINITION Sequence 1 from Patent WO0050581.
 ACCESSION AX034425
 VERSION AX034425.1 GI:10303130

KEYWORDS Neurospora crassa
 SOURCE Neurospora crassa
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 REFERENCE 1
 AUTHORS Cogoni, C. and Macino, G.
 TITLE Isolation and characterization of a n. Crassa silencing gene and
 JOURNAL Patent: WO 0050581-A 1 31-AUG-2000;
 UNIV ROMA (IT) ; COGONI CARLO (IT) ; MACINO GIUSEPPE (IT)
 FEATURES
 source Location/Qualifiers
 1..8045
 /organism="Neurospora crassa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5141"
 2447..6655
 /notes="unassigned protein product"
 /codon_start=1
 /protein_id="CAC10121.1"
 /db_xref="GI:10303131"
 /db_xref="REMBL:CA10121"
 /translation="NNPITFRKNSPVEIINRLNNDYNLGLQCVADTLTTPRRKEL
 ASEDDEFGHDKIYRALNLYWRKODSLNQANFIEAKAASNNVVKAHADPTLP
 WSKPPRAATAGQWALQVLLVLRNRPNNPTFGRTGLSGSLSGRTNT
 KRDEPANTVADPPKRSATGPIHGAALPLKFPDPVATGSKRPSLESENLCQ
 TXGAKGLSDNVAAAAPVPLASALDKVPTREHANTDPATGHRADQVDFSTSQ
 GTSGSVPSACHNQSTTQSSPEAPSPREKPVDAVFEAGHLIESFGKRTTKS
 HINDQPLSSSSQGETSFSTIYFSPSSGEGALPEPSRNGSLARSESARSOVQHAP
 VHAARLENWPKFKWLHAPLAVAWEVTRLFMHCKVDLEDLSGLKIDFESWATDV
 TDWTKLYRLDAFRGKFPPEKPNDFVTMTGNTFESKGSAAVLSAVLDYNDPNSPTA
 PLVYLKPLMFQCGQLTRRPGDFRFFELLPSPTSTSPSPVPVPVKQPGAVEEVIQ
 LMTWGHSLVGRWRAFFAKDAGYRKPLREFLRADPKPIIKERVHFFAETGTFP
 DVFKTSVLPAAEPQRTFVKYSOMLWLLQDNNTPWPHLKLFSRIQLGSLKTVAI
 MTLPEQIRHKKTDLLSPSGTEVNDGVGRMSRVAKRIRDLGLGDVPSVQGSFG
 SAKGMVVDVDDTGDDEWIEIYPSQRKWCDFVDKQRTLEVSVALESLKSLQNLQ
 LPVLEDRDKVKQRAIGLINDLQRFSEQKHALNRPVFRQWVYSESRAIRV
 SHGRVPLAGLPDSQBELTNLMNSGDFPKKYLQDIAWDLQKRCITLTKKLNLRV
 GSRAYIYVADFWGLNEEYHVFSGSKFDEEESFTLLDCDVLVARSFAHKKPDIQ
 RVAAPDLRSLKDVIIIPSTKGDPVLAKKLGGDYDGDMAWVNDPEIVDFGVNAEM
 PLBPDLSRYLKDITFKOLMASHGTGSAKEQTYDMIKSFHALQPNFLMGCTNY
 KERLCVNNNSVSKPALILSSLVGNLDVQSKGIVENEASWALREELGGLSLDPP
 MYKSDNLGRGETHIIIDLKESIAAPADKLEAFHNAMKAADTEDGAHPWDPDLA
 SYITFFKEISDKSRSSALLFTLKNRIIGEVKEYGRLVKVKMRSDSKDYPVYRNVQVY
 ERWCATTEAMDKSGNYSKVIRLLESLFLADRENMTWALLRSTAFKLYHKSKPKF
 VWQVAGRQLAYIKAQMTSRPGEAPALMTAFMYAGLMPDKKFTKQIVARLEGDGSFEP
 DPEVYEVLGDDDFDGIQTNGDY"

ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 8045
 Score: 7397.00 Matches: 1402
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-913-878A-2 (1-1402) X AX034425 (1-8045)

Qy 1 MethAsnProIleThrProArgLysArgAsnSerProValGluGluIleLeuAsnArgLeu 20
 Db 2447 ATGAACCCCTATTACTCTAGGAGAGGATAGCCCGCTCGAGGAATCATAAACCGGCTC 2506
 Qy 21 AsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThrThrLeuThrProHisArg 40
 Db 2507 AATACGACTACAACTTGGGCTCCAGTGTGTCCGACACACACTCTCACCCCCACCGC 2566
 Qy 41 ArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAla 60
 Db 2567 CGGAAGGAGCTGGCGGAGAGTACGAGGATTTCCGTGCGCATGACAGATCTACAGAGCC 2626
 Qy 61 LeuAsnPheLeuTyrTyrTrpArgLysAspSerLeuAsnGlnAlaGluAlaAsnPhePhe 80

Db 2627 CTGAACCTTCTCTACTGCGGAAGGATGACTCCCTGAAACCAGGCGAAGCAACTTCTTC 2686
 Qy 81 IleGluAlaLysAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThr 100
 Db 2687 ATCGAGGCCCAAGCTCGAGCTCGAAGCTGGGTGCCCAAGCCAGCCGACCTTGACAG 2746
 Qy 101 LeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnInTrpAlaLeuGln 120
 Db 2747 CTTGGTGTCTCAAGAAACCTCCCGCGCGCTACTGCGGCGCAACAATGGCATTCGAG 2806
 Qy 121 ThrValLeuLeuGluValLeuAsnArgPheMetProProAsnAsnThrProGlyArg 140
 Db 2807 ACTGTGTCTCGAGGTGTTAATAGGTTATGCCACCTCCCAATPACACACCAAGTCGA 2866
 Qy 141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
 Db 2867 ACGTTGGCAGAACTTAAGCGGCCAAGTGGCTGAGCGGCCCAACCTCTACCAACACC 2926
 Qy 161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
 Db 2927 AAACGCAAGGATGAGCCGCCCAATGTCACTTTTCGTGTATCGCCCAACGCTCGTTGACT 2986
 Qy 181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
 Db 2987 CGCTCTGCCACAGGCTCTCTATTACGGCGCGGCGATACCCCTAAAGTTCCCGATCCA 3046
 Qy 201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
 Db 3047 GTGAATACCGGTTCCAAACGACCATCTCTCGAGAGTGAGATCTCAATCACTGACCAACG 3106
 Qy 221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaIleProProValProIle 240
 Db 3107 CGGGCCCAAGGCAAGCTCTGATTAATGTTGCGGTGCGCGCGCCCGCTGCCATT 3166
 Qy 241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
 Db 3167 GCGAGCGCTTTCGACAAAGTACCGACTCGAAGCATGCGCAATACAGAGATGCCACGCG 3226
 Qy 261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
 Db 3227 ACAGTCAATAGCGAGCGACCAAGTGGATTCTTTTGATACATCTCAAGGCACCTTCCTAT 3286
 Qy 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
 Db 3287 GGTTCAGTGTCTTACGCGCTTTCGCGTCACAAATCAGACCACTTACCAGAGATGTTTTCAG 3346
 Qy 301 AlaProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGly 320
 Db 3347 GCTCCTCTTCAGCGCCAGAGAGAGCGGCTGTGGATGCCCGCTTTTGAGGCTGGA 3406
 Qy 321 HisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsnGlnPro 340
 Db 3407 CACTTGATTGAGTCTCTAGCAAGGAAGAAACAACCAAGTCCCATATAGATAACCGGCC 3466
 Qy 341 LeuSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
 Db 3467 CTTTCATCGTCTTCCAGGGTGAACCTTCGTTACGACCTTACTATGAGTCGTTTCCAAGT 3526
 Qy 361 SerGlyGlyGluGluAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
 Db 3527 TCCGGCGGCGAGGGCGCAATTCGCGCGAGTCGCTCAAATGGACTGGCTCGAGCGAA 3586
 Qy 381 GluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsn 400
 Db 3587 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCGCGCGGTGGTTGCGAGCTCGGCTGAGAAAT 3646
 Qy 401 IleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaIleTrpGluVal 420
 Db 3647 ATTTGGCCGGAATTTCCCAATGGCTTACACGAGCTCTCTCGCTGTGTGCATGGGAAGTT 3706
 Qy 421 ThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyr 440
 Db 3707 ACCAGACTCTTTATGCACTGCAAGTAGACTTTGGAAAGACGAGCGCTGGCGCTAAAGTAC 3766

Qy 441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeu 460
 Db 3767 GACCTTCTCGTGTCTACCGCGCGGATGCACAGATATCTGGAAGACTCTCTACCGGCTT 3826
 Qy 461 AspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThrAla 480
 Db 3827 GATGCTTTCGTTGTTAAACCTTTCAGAAAAAGCCGCCCAACGACGTGTTCGTGACCGCA 3886
 Qy 481 MetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyr 500
 Db 3887 ATGACGGGCACTTTGAGAGCAAGGTAGTGGCGTGTCTCTCTCTGCTGTCTAGACTAC 3946
 Qy 501 AsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPhe 520
 Db 3947 AATCCCGACAACTCGCTACTCGGCCCTTTTACCTTGTGAAGTGAAGCGGCTCATGTTTC 4006
 Qy 521 GluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPheGluLeuLeuLeu 540
 Db 4007 GACAGGGCTGTGCATCTACCCGTCGGTTCGGTTCCTGATAGGTTTTTCGAGATCCCTATA 4066
 Qy 541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560
 Db 4067 CGCTCGCTACGAGCACCAAGTGTACCGCGGTGTCTCAGCAAAACCAACACGAGTGCG 4126
 Qy 561 ValGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrp 580
 Db 4127 GTCGAAGAAGTCACTCAGTGGCTCAGATGGGGCAACATCTCTTGTAGGCGCCCAATGG 4186
 Qy 581 ArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArg 600
 Db 4187 CGCGCTTCTTCGCCAAAGATGCCGATACAGAAACCTCTCAGGAGTTCCAGCTCCGC 4246
 Qy 601 AlaGluAspProLysProIleLysGluArgValHisPheAlaGluThrGlyIle 620
 Db 4247 GCGAGAGCCCGAAACCCATCATCAAGGAGAGAGTCCACTTCTTCCCGAGACCGCAAT 4306
 Qy 621 ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
 Db 4307 ACGTTCGACCTGATGTGTTCAAGACGAGATCTGTGTTCCGCGCAGAGAAACCTGTAGAG 4366
 Qy 641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsn 660
 Db 4367 CAACGAGCCGAGTCAAAAGTACTGAATGCTGAGTGGCTCTGCAACTCGCAACAAC 4426
 Qy 661 ThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyr 680
 Db 4427 ACTTGGCAGCGCCACCTCAAGTTGTTCTCCGTTATCCAGTCCGTTCTGAGTAAGACATAT 4486
 Qy 681 AlaIleMetThrLeuGluProHisGlnIleArgHisIleLysThrAspLeuLeuSerPro 700
 Db 4487 GCCATTTATGACATTTGGAGCTCCACAGATCAGACCAACCAAGACCCGATCTCTTTCACCT 4546
 Qy 701 SerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLys 720
 Db 4547 TCAGGCACTGCGAGTGTATGATGATGAGGTGAGCGCATCTCGCGAGCGTGGCCAG 4606
 Qy 721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
 Db 4607 AGGATACCGATGTTCTCGGTTTTCGGTGTATGTCCTCTGCTGTCAAGGCGGTTTGGT 4666
 Qy 741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
 Db 4667 TCGGCCAAGGAAATGTGGTTATCGAGCTTACGACACAGCGCATGAGGATTCGATTCGAG 4726
 Qy 761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
 Db 4727 ACATACCGTCCCGAGCGCAAGTGGAAATGCGATTCGTTGATAAACATCAACGATCCCTC 4786
 Qy 781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 800
 Db 4787 GAAGTCCGAGCGTGGCTTCTGAAGTCAAGTCAAGTGGTCTCAACCTACAGCTGTACCT 4846

QY 801 ValLeuGluAspArgAlaAspIysValIysMetArgGlnAlaIleGlyAspArgLeu 820
 DB 4847 GTCTCTGGAAGATAGACCCAGGACCAAGGTGAAGATCGCGCAGCAATCGGTACCGCTTT 4906
 QY 821 IleAsnAspLeuGlnArgGlnPheSerGluGlnIysHisAlaLeuAsnArgProValGlu 840
 DB 4907 ATCAACGATTTGCAACGACAGATTCAGCGAGCAAAAGCATGCTTTGAAATCGCCCATGGAA 4966
 QY 841 PheArgGlnTrpValTyrGluSerTyrSerArgAlaThrArgValSerHisGlyArg 860
 DB 4967 TTTTCGCAATGGGTTCACGAGAGTTATTCAGTTCGCGCAACTCGAGTCAGCCACGCGCGT 5026
 QY 861 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 880
 DB 5027 GTGCCCTTTCTGCTGGGTACTCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAA 5086
 QY 881 SerGlyPheAspProIysGlnIysTyrLeuGlnAspIleAlaTrpAspLeuGlnIys 900
 DB 5087 AGTGGGTTCGATCCCAAGAGCAAAAGTACTTTGCAAGACATCGCTGGGATCTTCAAAAG 5146
 QY 901 ArgLysCysAspPheLeuIysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
 DB 5147 CGGAATGTGACACGTTGAAGTCCAAAGCTGAACATCCGTCGTCGATCAGCATACATT 5206
 QY 921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
 DB 5207 TACATGATTGCCGATTTCTGGGTGTGCTTGAGGAAATGAGGTTCATGTGCGATTCTCC 5266
 QY 941 SerLysPheArgAspGluGlnGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
 DB 5267 TCAGATTCAGACAGAGAGAGTCTTTTACACTCTATCGGATCTGTGATGCTCTCGTG 5326
 QY 961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
 DB 5327 GCGCGATCCCGACCCCATTTCCCTAGTGATATCCAAACGGGTTCCGAGAGTCTTCAAGCCA 5386
 QY 981 GluLeuHisSerLeuIysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
 DB 5387 GAGCTCCACAGTCTCAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGTTCCT 5446
 QY 1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
 DB 5447 AAGAAGCTATCTGTGTGAGACTACGACCGCGATATGGCCCTGGTCTGTCTGGGATCCGGAG 5506
 QY 1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
 DB 5507 ATCGTCGATGGTTTCGCTCAATCGGAAATGCCCTCTGGAGCCGACCTGTCTAGGTACCTA 5566
 QY 1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
 DB 5567 AAGAAGGACAAACGACTTTCAAACAACTTATGGCTCTACACGGCACCGGCTCAGCGGCC 5626
 QY 1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
 DB 5627 AAAGAGCAGACTACATACGATATGATCCAGAAAGAGCTTCCATTTCCGCTCGCAGCCCAAC 5686
 QY 1081 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 1100
 DB 5687 TTCTTGGGATGTGCACTAACTACAAAGAAAGGCTCTGTATCATCAACATAGTGTCT 5746
 QY 1101 AsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 1120
 DB 5747 AACCAAGCGCGCCATCATCTTACTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAA 5806
 QY 1121 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyAla 1140
 DB 5807 GGTATTGTCTTTAACCAAGCAAGCTGGGTCAATTCGTFAGGAACTGCTTGGCGGTGCA 5866
 QY 1141 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThr 1160
 DB 5867 TTGTCTCTCTGACCCCAATGTACAAAGAGCAGCTTGGCTCGGCGCGGAGAGCCTACC 5926
 QY 1161 HisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 1180

DB 5927 CACATTATTGACFACCTGAAATTTCTCCATCGCCAGCGCTCGGATTGACAAAGAACCTGGAA 5986
 QY 1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
 DB 5987 GCCTTCCACAATGCCATGAAAGCGGCAAGGATACAGAAAGACGGCGCTCACTTTTGGAT 6046
 QY 1201 ProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSer 1220
 DB 6047 CCGGATCTCCCTTCCACTACAGCTTCTTCAAGGAGATTAGCAGCAAGTCGCGATCGTCC 6106
 QY 1221 AlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg 1240
 DB 6107 GCACTGCTATTACGACTCTGAAAGACCGTATCGCGAAGTCGAGAAAGAAATATGCGAGG 6166
 QY 1241 LeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGln 1260
 DB 6167 TTGTCTCAAAACCAAGGAGATGAGAGACGACGAGGACCCCTTACCCTGTCCGCGTCAACCAG 6226
 QY 1261 ValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyr 1280
 DB 6227 GTTTATGAAAATGGTGGCCATCACGCTGAGCGATGACAAATCCGAGCAAAATAT 6286
 QY 1281 AspSerLysValIleArgLeuLeuLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
 DB 6287 GATTCTAAGGTGATCAGGTTGCTGGAGCTGTCTTCTCCGAGACCGTGAGATGATACA 6346
 QY 1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
 DB 6347 TGGCATTTGCTGAGGGCTAGCAGCGCTTTCAGCTGTACTACCAAGAGGCCCAAGTTC 6406
 QY 1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
 DB 6407 GTGTGGCAGATGGCGGCGACACAGCTCGCGTACATTAAGCGCGAGATGACGAGCAGACC 6466
 QY 1341 GlyGluGlyValProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
 DB 6467 GGTGAAGCGCCCGCGCTTGATGACCGCTTCATGTATCGCGCTTGATCGCGATAG 6526
 QY 1361 LysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspPro 1380
 DB 6527 AAGTTTACGAAGCAGTATGTGGCCAGCGCTGAGGCGGATGGATCGGAGTACCTGTATCCG 6586
 QY 1381 GluValTyrGluValLeuGlyAspAspPheAspGlyIleGlyPheThrGlyAsnGly 1400
 DB 6587 GAGTCTATGAGTGTGGCGCATGATGATTTTGATGGAATTTGTTTTCACAGGGAATGGC 6546
 QY 1401 AspTyr 1402
 DB 6647 GATTAT 6652
 RESULT 3
 AP004357/c
 LOCUS
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 BAC clone: B1074C08.
 VERSION AP004357.4 GI:21328130
 ACCESSION AP004357.4
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzae; Oryza.
 REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
 Karayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y.,
 Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
 Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
 Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y.,
 Igonuma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
 Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,
 Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,

Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshioka, S., Yoshinaka, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y., Yano, M., Jiang, J., and Gojobori, T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 165701)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission

Submitted (19-NOV-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail:tsaaki@nias.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

On Jun 6, 2002 this sequence version replaced gi:20975426.

Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTX2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against

NCBI NonRedundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db)

and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database using BLASTP2.0. ESTs represent

the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level

such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering

almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program

is classified as a 'hypothetical' protein.

The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. This sequence of B1074C08 clone has an overlap with B1147B04

clone (DBJ:AP004641) at the position 1 to 108,335 of 5' end and an
overlap with B1125H01 clone (DBJ: AP003370) at the position

115,348 to 165,701 of 3' end. Detailed information on overlap and
assembly quality together with annotation of this entry is

available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

Location/Qualifiers

FEATURES

source

1. .165701
/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosomes="1"

/clone="B1074C08"

/complement(join(1956..2373,2716..2783))

/gene="B1074C08.1"

/complement(join(1956..2373,2716..2783))

/gene="B1074C08.1"

/note="contains EST D22801 (C1352)

/unknown protein"

/codon_start=1

/protein_id="BAC00710.1"

/db_xref="GI:21328131"

/translation="WDSTLTQKATATATLITSVIRINBELGTSNKKPKGASFO
ANESVLTWFIWAFDSDSMVIVQCHTSLTLEKSYFCYCYLIHFLLASMTYKPT
PFWHTQPKKNQNFRRSDPTNEATNRTEEDRGMDLEDVGGQSCQRSMYTORWM
M"

join(7311..7428,7640..7776)

/gene="B1074C08.2"

join(7311..7428,7640..7776)

/gene="B1074C08.2"

/note="contains EST AU069673 (C0828)

/unknown protein"

/codon_start=1

/protein_id="BAC00711.1"

/db_xref="GI:21328132"

/translation="MHCHCYRCNVRSYQAQQPPVRACVAAQPRWKVHNTNTRPAARM
KLIGSGSGGQQRGIARALKEHAKLYIIRRCVWMLLR"
join(9087..9169,9577..9836,9998..10158,10204..10298,
10342..10570)
/gene="B1074C08.3"
join(9087..9169,9577..9836,9998..10158,10204..10298,
10342..10570)
/gene="B1074C08.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC00712.1"
/db_xref="GI:21328133"
/translation="MARQCLDRVGRDGGAAVAKREREREVTDMGPPVGPFFHLP
PFPKATATPAPLPAAPRLSRTPAPSPRRRLPRLPSLPAHARAPKAGPFLNPP
PTLSLSIPIHVGPSPSRAVIAFGACSRASISTPLSKGAIGRHSPCAGEPFRRR
RRWFSLTVVLEVRSLPPFPGRSPSLFAGVFPHAGRI PRAEPRRRSRHGLRPTLPP
PAPVTTRAASAASHGLVGEAPSPAPAAAPRAVSARGSRQSGCGPAVDHPRGPGP"
join(16765..117574,19784..20080)
/gene="B1074C08.4"
join(16765..117574,19784..20080)
/gene="B1074C08.4"
/codon_start=1
/product="putative DNA damage repair protein"
/protein_id="BAC00713.1"
/db_xref="GI:21328134"
/translation="MLGGLXGLDLPSPSSADDDKPAAGHSSAAKMAPPTLRKPPATP
APEDYLRNRRAPKAPAAQPPPTLPIETTSFQPALVAVQSTVMEYDPARP
DYEDYKRLKRAEAREKLEEREREERERELREGRDNLNIGEEAWK
AAGSGAAPSPSPHGDGFAIGNSSSSGLGAGQMTAAQRMAMWGKEGGGLG
KQCGITAPVANKTDRRGVIVDENSCKOEKPKSVNFDGPTFVLLRNMYGPGEV
DLEERVASECCKYGVILVLIPEITQADFPAAEAVRIILFERAEATKAMIDLEG
RFGGRVVRATFFDESEFGKQLAPMGVAGFD"
join(22181..22363,22471..22640,22731..22803,22919..22993,
23119..23329,23337..23382,23475..23538,24200..24302,
24389..24484,24576..24605,24824..24923,25151..25235,
25355..25478)
/gene="B1074C08.5"
join(22181..22363,22471..22640,22731..22803,22919..22993,
23119..23329,23337..23382,23475..23538,24200..24302,
24389..24484,24576..24605,24824..24923,25151..25235,
25355..25478)
/note="contains ESTs
C26310 (C12073), D15391 (C0569), AU100694 (C12073)
unknown protein"

/codon_start=1
/protein_id="BAC00714.1"
/db_xref="GI:21328135"
/translation="MAPGTLAPKRRKAEASPSFSPMGSDSGYSDLDHABES
FYSGSGSDRQVSSNDDDEEEEREMDEEDDDDDDEEMNEDDEDEGE
MNELEKEYRTLQTNQNILETQKRRDDVSKQAVKQKQVLDKALEMRFLQKAFS
TNSKLPEKPIRSMFCNORIEOAYLDLNSKOTLGSMWELQELERNRATKDVTD
TNSLNGEDDEWSEVOKLOEITPFRNSEIDKQKQTQVTTGAALKKGLHAFQND
ISQVTSYMRDSESRMINRHLKSLTGVFGEVGEHENKKNTEGDELDVDSSEFY
QQLKBEFLSCDAGASESAYALKQKQKRLVDRSRASKIRIVHVEKIANFAP
VNVPIPPMAPKLFENLFGMGNQKSTTA"

join(26433..26525,30039..30091,31996..32168,32606..32724)
/gene="B1074C08.6"
join(26433..26525,30039..30091,31996..32168,32606..32724)
/gene="B1074C08.6"
/note="hypothetical protein"

/codon_start=1
/protein_id="BAC00715.1"
/db_xref="GI:21328136"
/translation="MEIHRTKLIDAAAQAVISDRRVVYVTCIMKPSSSSVVDTTSP
LSQHRGVTPLLAVRSRGLAHFFHFSDQPSLLTPAKSAMWMLTPDVQATLHAULTQ
LRVQSTTTTEKLIGAGWEPALGAGFPNRRHGGDDTDFQSCRFP"

36236..36550
/gene="B1074C08.7"
36236..36550
/gene="B1074C08.7"
/note="hypothetical protein"
/codon_start=1

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

/protein_id="BAC00716.1"
/db_xref="GI:21328137"
/translation="MAAALQIWRRRRGAFAAAPPATRRRCRRRLPSRYHPHPPPPPL
LFGCRYSYLVGRRCRRRLPSCHHPSPSPPLPLFRHSRPLDLVDAASTAGPPP
RV"
complement(38211..39233)
/gene="B1074C08.8"
complement(38211..39233)
/gene="B1074C08.8"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC00717.1"
/db_xref="GI:21328138"
/translation="MDWASLPADLLFSISSHLREPEDFVRPEAVCPQWRAAVSHKEHA
FPQPMASRMLEDSENVLYSLSTKTIKVHPDMKSRIRASGSHLAIADKDD
DLSAVLINESEKTHALPRLPFFHDNGAGHWITGEGVITVYVNNWSENWALWYRG
GITMKWAVPRKULRSHYRLKLAAYGDDMEMHLLDGLGDNDSDSVLLQETKKVE
LGGCWPREDVEKTPCHHEWFSLYRNVEQEIIPVHDIGNVMVQSRDSCRTYMI
PASRDFAALGSRNAPYILWKQFDAGSYNALFKKLASEVLTFFVKRLPEDWKLSDWF
MPSLKY"
LTR
42943..43587
/note="5' LTR"
join(44015..44093,44667..44836)
/gene="B1074C08.9"
join(44015..44093,44667..44836)
/gene="B1074C08.9"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC00718.1"
/db_xref="GI:21328139"
/translation="MPFRKLEAAPRRVPRRASLPCCAAAEPEGVADLSEGEDPODYE
QKKSUFFHVPNCBIIILFTNKIACNDHEPTCDYWP"
join(47189..47443,47555..47728)
/gene="B1074C08.10"

Alignment Scores:

Pred. No.: 0.000141 Length: 165701
Score: 348.50 Matches: 242
Percent Similarity: 35.55% Conservative: 154
Best Local Similarity: 21.72% Mismatches: 383
Query Match: 4.71% Indels: 337
DB: 8 Gaps: 51

US-09-913-878a-2 (1-1402) x AP004357 (1-165701)

Qy 102 ProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThr 121
Dy |||||
Db CCTGAGGGGAGAGCCCTCGCGCTCGCGCGCGGACAG-----CTCGCTCGG 113613
Qy 122 ValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrPro----- 138
Dy |||||
Db CGTGTCCCGCCGAGCAGATCCGCGCTTCGGTTCGCCGAGCGCGCTCGAGTCCG 113553
Qy 139 GlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArg-----Pro 155
Dy |||||
Db GGGCGTCCCGTCCCGCGCGCGCT-TCCTCGCGCGCTGGCGCGCGGCGGACGCGGCTCG 113494
Qy 156 ThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspPro 175
Dy |||||
Db ACCTTCCTCGTGACCCCTTCGAGCGGTGCTGCGCGCTCGCTGTTCACCCGCGACCGCT 113434
Qy 176 LysArgSerLeuThrArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeu 195
Dy |||||
Db TCACGTTC--CGGGCTTCGCGAGGTCCGCCCATAC-----GCTCGCAGCTCA 113386
Qy 196 LysPheProAspProValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeu 215
Dy |||||
Db AGCTGGAGTTCCTCGCGCGAGCTCTCGGAGG----- 113353
Qy 216 AsnGlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAla 235
Dy |||||
Db TCCGCTGTATCC-----GCTCGACTGCTCGCTGCTCCTCGCGCTCG 113311

Qy 236 ProProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThr 255
Dy |||||
Db 113310 CCGCG-----CGCGCTGGTGCATACC 113287
Qy 256 ArgAspProThrAlaThrGlyHisArgAlaAspGlnValAspSerPheAspThrSer 275
Dy |||||
Db 113286 GCACCGCGGACGAGCTTCCACGAGCGGTCCCTTCGACCTGCTCGACGACGAGACC 113227
Qy 276 GlnGlyThrSerTyrGlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThr 295
Dy |||||
Db 113226 CGTGATACGACGACGACATCACCCGAGCGCGCATCGCGCGGTGCGGGGTGTACA 113167
Qy 296 GlnSerSerPheGluAlaProProSerGlnProArgGluLysArgProValAspAlaThr 315
Dy |||||
Db 113166 GG-ATCTCTCTCGCGG---CGATTCTGCCCAAGATGACCGCGCTGGATACATG 113111
Qy 316 ValPheGluAlaGlyHisLeuIleGlu-----SerProSerLysGlyArg 330
Dy |||||
Db 113110 AGGAGAGGAGGTGGCGATCGTTCGATTCGCGCGCGATGCGGGCGCGCGGGGCTC 113051
Qy 331 ThrThrLysSerHisIleAsp---AsnGlnProLeuSerSerSerGlnGlyGluThr 349
Dy |||||
Db 113050 ACAGTGCAGCAGCTGGAGTTCGGGAGCCATGACG-----GACGTC 113006
Qy 350 SerPheSerThrTyrTyrGlu-----SerPhePro-----Ser 360
Dy |||||
Db 113005 TTCTTCTGCTGCAGCAGCGCGGAGGCTCAAGTTCCTCGCTCTTTCATGTGTAACGG 112946
Qy 361 SerGlyGlyGluGlyAlaIleProGlu-----ProSerArgSerAsnGlyLeu 376
Dy |||||
Db 112945 TTGGTGCAAGGAATCAACACGACACCACTCAGCTCAATTCACGCTGCTG 112886
Qy 377 AlaArgSerGluGluSerAlaArgSerGlnValGlnValHisAlaProValAlaAla 396
Dy |||||
Db 112885 GGGAGAGCGGAGGAGAAC-----GTCAATGTGGCT 112856
Qy 397 ArgLeuArgAsnIleTrp---ProLysPheProLysTrp-----LeuHis 410
Dy |||||
Db 112855 GCATTGAGGATTTCTGGGGGACAAGTTCCTAGTGTTCATCGCTGGGAGGCTGAAG 112796
Qy 411 GluAlaProLeuAlaValAla----- 417
Dy |||||
Db 112795 AAGCACTGAATCGGTGGCGGAGAACCCCAACTTCTCTGCAGCAAGTTCGGGATGAC 112736
Qy 418 ---TrpGluValThrArgLeuPheMet-----HisCys---LysValAsp 430
Dy |||||
Db 112735 CACGCGAGGTGCGGAGCTGGTGATCACGCCACGAGGCTTATTGTCGCTCCAGAA 112676
Qy 431 LeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrpSerThrAlaArgAspVal 450
Dy |||||
Db 112675 GTGGAGCGCTCAACCGTGTCTTCGGCAT----- 112646
Qy 451 ThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPheProGlu 470
Dy |||||
Db 112645 -----TACCATGAG 112637
Qy 471 LysProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGlySer 490
Dy |||||
Db 112636 GTGGCTACAGGTTTTCAGGGTCACTTTATGACGAGCGGTATGAC----- 112589
Qy 491 AlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAlaProLeu 510
Dy |||||
Db 112588 -----GTGCTCAACACAATGTCTCAACTCTCTCACTGCACCAAT 112547
Qy 511 TyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPhe 530
Dy |||||
Db 112546 -----GTCAAGACTTGTATG----- 112532
Qy 531 GlyProAspArgPhePheGluIleLeuIleProSerProThrSerThrSerProSerVal 550
Dy |||||
Db 112531 -----TCGAATTTTTCAG----- 112517
Qy 551 ProProValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMet 570

| | | | |
|----|--------|--|--------|
| Db | 111706 | TTGAAAGAACGAGGATATTGTGCGCAAGGGAAGGTGGTGTATGGCGCTGTCTTGATGAG | 11164 |
| Qy | 927 | TrpGlyValLeuGluGlu-----AenGlu | 934 |
| Db | 111646 | CTTGGGTTCTTGAGCAAGGCGAGTGTATTTCGGGCAACAGATTCCATCATTTGAATAGT | 111587 |
| Qy | 935 | ValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSer | 954 |
| Db | 111586 | TATTTTGTTAAGCANTGGTCAAGATTTCATCAACAGATAAAAAACAGAGTCATTTTG | 111527 |
| Qy | 955 | AspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVal | 974 |
| Db | 111526 | GGT--ACTGTGGTAATAGCAAGAAATCCCTGTCTTCATCCAGGGGATGTCGGCATCCTT | 111470 |
| Qy | 975 | ArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLys | 994 |
| Db | 111469 | GAACAGTTGATGTCGCCGAACATCATCTGTTGATTGTTGGTGTCTCCCGAGAAA | 111410 |
| Qy | 995 | GlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrp | 1014 |
| Db | 111409 | GGTGAGAGGCCACATGCTAACGAGGCATCTGGAGCGCATCTTGATGGGATCTCTACTTT | 111350 |
| Qy | 1015 | ValCysTrpAspProGluIleValAsp----- | 1023 |
| Db | 111349 | GTGACATGGGATGAGAAATCTTATACCTCAGGCAAGAGACTGGAACCTTATGGACTAC | 111290 |
| Qy | 1024 | -----GlyPheValAsn | 1027 |
| Db | 111289 | TCCCCCCTGAAGCAAAACAACCTTCGGGCCAAGTATCTCAACATGTAAGTTCATAACT | 111230 |
| Qy | 1028 | AlaGluMetPro-----LeuGluProAspLeuSerArgTyrLeuLys | 1041 |
| Db | 111229 | CTTGAGATGTCAGCAAGTGTCCCAATTTTGTTCATACATCATGCTCTTTTATAC | 111170 |
| Qy | 1042 | LysAspLysThrThrPhe----- | 1047 |
| Db | 111169 | GATGGATAACCTCTCTTCATTGGAAGCATATATGGTTCCTTTACAGAACTGAGATAA | 111110 |
| Qy | 1048 | -----LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr | 1064 |
| Db | 111109 | AGAAATTAAGCGCAAGTATGTGTCGGGCAC---ACTGCACACCCAGCAAAAAACGGGAAT | 111053 |
| Qy | 1065 | ThrTyr-----AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe | 1081 |
| Db | 111052 | AGCTTTAGATCTCAGATATATTAACGAATCA-----AAATTAAAGTCCCAACACAG | 111002 |
| Qy | 1082 | LeuGly-----MetCysThrAsnTyrLysGluArgLeu | 1092 |
| Db | 111001 | AGAGTCTCTCTACATTTGTAACCTTATAGGCATCATCTC | 110960 |

NTA011576

| | | | | | |
|------------|--|------------|------|--------|-----------------|
| NTA011576 | NTA011576 | 3505 bp | mRNA | linear | PLN 25-FEB-2003 |
| LOCUS | Nicotiana tabacum mRNA for RNA-directed RNA polymerase. | | | | |
| DEFINITION | | | | | |
| ACCESSION | AJ011576 | | | | |
| VERSION | AJ011576.1 | GI:4138281 | | | |
| KEYWORDS | grp gene; RNA-directed RNA polymerase. | | | | |
| SOURCE | Nicotiana tabacum (common tobacco) | | | | |
| ORGANISM | Nicotiana tabacum | | | | |
| REFERENCE | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Nicotiana. | | | | |
| AUTHORS | 1 Schiebel,W., Pelissier,T., Riedel,L., Thalmeir,S., Schiebel,R., Kempe,D., Lottspeich,F., Saenger,H.L. and Wassenegeger,M. Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Tomato | | | | |
| TITLE | Unpublished | | | | |
| JOURNAL | 2 (bases 1 to 3505) | | | | |
| REFERENCE | Wassenegeger,M. | | | | |
| AUTHORS | Direct Submission | | | | |
| TITLE | Submitted (16-OCT-1998) Wassenegeger M., Max-Planck-Institut of Biochemistry Virolidressarch, Am Klopferspitz 18A, 82152 | | | | |
| JOURNAL | | | | | |

FEATURES
source

Martinried, GERMANY
Location/Qualifiers
1. .3505
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Petit Havana SRI"
/db_xref="taxon:4097"
/tissue_type="leaves"
1. .3505
/gene="RdRP"
20. .3370
/gene="RdRP"
/codon_start=1
/evidence=experimental
/product="RNA-directed RNA polymerase"
/protein_id="CAA09697.1"
/db_xref="GI:4138282"
/db_xref="GOA:Q9ZRY7"
/db_xref="SPTREMBL:Q9ZRY7"
/translation="MGKTIQVPPPLYSABAVKFLNHTGNGTVVALEVKQSGGR
RPAKQVQFANNKSAEFLIDASKGLYGSYLKAWEMKTDIVQPTVVEHVDGVTLPF
GCQIDKFAVLGSKDVSIGKIGLAKIYFSLSHASADYKQLQSYENIMQVVLPRPV
QNAQLLQGFAPRIYKLEDCSYFKEPTDDQWRTDFTFSWIGLSSSLCEFR
NGVQLPNSSEFFYKESNMQFIQTGTFSSOKLALPVLVHLPGLIEHLYKILFKI
SSLIQGFPPGLAFNFPQLVPRRNIACTEHALEKLYLKECCYDVPWLTQYD
EYLKGRQLPSPPIITLDGLVYVRVVTVPCKVFCGPEVNSVRVLRNVEDINNEL
RYSFVDEBEKTHSDTLPRAGTNGRTDIYERILSLRNGFTIIGDKRFFLAFSS
QURDSVWFEASRPGLTANDIPTWMDRQIRNVAKYARLGQSGSSRETLVSGRHE
VVIDPVACSLHGTNYIFSDGIGKISADFAHRVTIKCGLQYTPSPQIRYGGYKGVH
VPIYSMSKLRSKMLKYESNNIKLDVGNRSKYQCIYLRQVILLSTLIGLVIDVLEQ
KQNEAVLQDAITLHDSLKAQEALELMSPEENTNLIKMLNCGMPDAPFFLSMMLQTP
RASKLQDTRITRIPIGRNTWVGCLDESRTLEYQGVQFVSGAGRSQFFFEESHPPND
SGSANDFLTKGNVAVKPNCLHPGDIIRLVAVDPALHVMVDVCPVFPQKGRPHNE
CSGSLDGLDGIYFCVDPDPLIPRQVQSMYDTPAPTQLDHVDVIEVEVEYFTNYIIND
SLGILANAVFADREPDMAMDPCQLAQLPSIAVDPPKTVGPVPAEIPSLQRPKEYPD
FHEKPKPYENKGVIGLFOKVKNTQOASSIAFTEDVARKVDSMDIVDGPEDYI
DEAFYKYSNKGILNMDYIGIKTEIELSGCINKASKTFDRRKDAEAIQVAVRCLR
KQAWFKRSRSDIDMLAKASAWHYVHTHTWGLYNEGLKRDHFISFPWCYVDQLIQ
IKKAKRKRPVPELSGLSKRLVIM"

ORIGIN

Alignment Scores:
Pred. No.: 7,77e-06 Length: 3505
Score: 328.00 Matches: 210
Percent Similarity: 34.06% Conservative: 151
Best Local Similarity: 19.81% Mismatches: 385
Query Match: 4.43% Indels: 314
DB: 8 Gaps: 42

US-09-913-878a-2 (1-1402) x NTA011576 (1-3505)

Qy 441 AspProSerThrAlaArgAspValThrAspIleThrIleThrLeu----- 457
Db 596 GATGATCAGTGGTGAAGGACACAGATTTCACTCCATCTGGATAGGGCTATCTTCTAGC 655
Qy 458 -----TyrArgLeuAspAlaPheArgGlyLysProPhePro 469
Db 656 TTATGTTGAGTTCGTANTGTTGTTCACTTCCAAATTCAGCGAAAGTTTTTTTAC 715
Qy 470 GluLysProAsnAspAlpHeVal-----ThrAlaValThrGlyAsnPheGluSer 487
Db 716 TATAAAGAAAGATGACCAATTTATTTACAGACTGGTGTTCACCTTTTCATCTCTCAA 775
Qy 488 LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThr 507
Db 776 AAATGGCTCTGGTTCCCATTTGTC-----CACCTTCTGAGGGAATGAGTTG 823
Qy 508 AlaProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCys-----Arg 525
Db 824 CCCTACAAGATATTTGTTCAAAATTAGTTCCTTGATA---CAACATGGATGCTTCTCTGGG 880
Qy 526 LeuThrArgArgPheGlyProAspArgPheGluIleLeuIleProSerProThrSer 545
Db -----

Db 881 CTAGCTTTAAATTTT-----ACCTTTTCCATTAGTTGACCTCGAAGGAGAAAT 931
Qy 546 ThrSerProSerValProValValSerLysGlnProGlyAlaValGluVal----- 564
Db 932 -----ATTGCATGCATTGAACACGCTTGGAGAAACTGTATC 967
Qy 565 -----IleGlnIleThr-----MetGlyGlnHisSer 574
Db 968 TATTAAAGAGTGTGTATTATGATCCAGTGAGTGGCTAACTGAGCAGTATGATGAGTAC 1027
Qy 575 LeuValGlyArgGln----- 579
Db 1028 CTCAGGGCAGACAACTTCCAAAATCTCGCCCATCACTTTAGATGATGATGGTGTAT 1087
Qy 580 -----TrpArgAlaPhePheAla-----LysAspAlaGly 589
Db 1088 GTAAGAAGGTCGTAGTAACACCATCAAGATTATTTTGTGTCCAGAGGTAACGTA 1147
Qy 590 TyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProLysLeuLys 609
Db 1148 TCCAATCGGTTCTCGCAATTAT-----TCTGAAGACATAAATAACTTTCTTCA 1198
Qy 610 -----GluArgValHisPhePhe-----Ala 616
Db 1199 GTTCTTTTGTGATGAGGAGTGGGAGAAATACATCTACAGATTATATCCAGAGCT 1258
Qy 617 GluThrGlyIleThrPheArgProAspValPheLys-----ThrArgSer 631
Db 1259 AGTACTGGAAATGTTACCAGGACAGACATCTATGAGAGATCCTCACTCTCGAAT 1318
Qy 632 ValValProAlaGluProValGluGlnArgThrGluPheLysValSerGlnMetLeu 651
Db 1319 GGCTTTATAATGGTGATAAAGATTGAATTTCTTGCATTTTTCATCGAGCCAGTTGCGG 1378
Qy 652 Asp-----Trp-----LeuLeuGlnLeuAspAsnAsnThr 661
Db 1379 GATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAATGATATAAGAACA 1438
Qy 662 TrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeuGlyLeu 676
Db 1439 TGGATGGGTGATTTTCGCAAAATCAGGAATGCGAAATATATCTGCCAGACTTGGCCAA 1498
Qy 677 SerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisIleThrAsp 696
Db 1499 TCITTTGGTTCCTCCAGAGAACTTTG-----AGTGTGTAGGATGAAGTTGAA 1549
Qy 697 LeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db 1550 GTTATTCCTGATGTAGCTTGTTCGCTTCATGGAACCAACTACATCTTTTTCAGATGGAAT 1609
Qy 712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
Db 1610 GGAATAATATCTGCTGACTTTGCTCATAGATTACCAATAAATGTGGCTTCAATATCT 1669
Qy 732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
Db 1670 CCATCTTCTTTTCAGATTCGTTATGTTGATATAAAGTGTGTGTGTGTGTGTGTGTGT 1723
Qy 752 AspThrGlyAspGluAspTyrIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
Db 1724 -----CCATATTCATCAATGAAG----- 1741
Qy 772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db 1742 -----TATCTTTGAGAAAGAGCATGTTGAAATATGAATCA 1777
Qy 792 AlaGlyLeuAsnLeuGlnLeuLeu-----ProValLeuGluAspArg 805
Db 1778 AACAACTAAGATTAGATGCTCTGGATGGAGCAAAATATCAGCTTCTTATCTTAACTGT 1837
Qy 806 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsnLeuGln 825
Db 1838 CAACTGGTTACTCTCTTCTACT 1897

```
QY 826 ArgInPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpVal 845
   |||
   |||
   |||
Db 1898 AATGAAGCAGTAGATCAGCTGATCTATTTTGCATGATTTCTTGAAGGCACAGGAAGCT 1957
   |||
   |||
   |||
QY 846 TyrGluSerTyrSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 865
   |||
   |||
   |||
Db 1958 TTAGATTGATGCTT----- 1972
   |||
   |||
   |||
QY 866 GlyLeuProAspSerGlnGluGluThrIleuAsnPheLeuMetAsnSerGlyPheAspPro 885
   |||
   |||
   |||
Db 1973 -----CCTGAGAGAAATACCTAATATCTCAAGAGATGCTAAATTTGCTGTTATATGCTT 2026
   |||
   |||
   |||
QY 886 LysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThr 905
   |||
   |||
   |||
Db 2027 GATGCTGAGCGCTTCCTTCATGATGCTGCAAACTTTCCGGGCTTCGAAGTTCCTCGAT 2086
   |||
   |||
   |||
QY 906 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 925
   |||
   |||
   |||
Db 2087 TTGGGACTAGACACAGAATATTTATTCCAAATGGTAGACGATGATGGGATGCTTGAT 2146
   |||
   |||
   |||
QY 926 PheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSer----- 941
   |||
   |||
   |||
Db 2147 GAATCCAGAACCCCTGGAATATGGTCAGGTGTTTCTCAGTTTCTGCTGCTGGACGTAGA 2206
   |||
   |||
   |||
QY 942 LysPheArgAspGluGluLysPhe-----ThrLeuLeuSerAspCysAsp----- 957
   |||
   |||
   |||
Db 2207 CAGTTTTTGGAGATCGCATCTTTAATGACAGTGGATCCGCAACTGATGTTTCATT 2266
   |||
   |||
   |||
QY 958 -----ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 973
   |||
   |||
   |||
Db 2267 CTCAGGGAACGCTGCTGTTGCAAAAATCCATGCTTGATCTCTGCTGATATTCGTTT 2326
   |||
   |||
   |||
QY 974 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 993
   |||
   |||
   |||
Db 2327 TTAAGGCTGTGATGTTCCAGCTTTCACACATGTTGATGTTGTTGTTATTTCCCCAG 2386
   |||
   |||
   |||
QY 994 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 1013
   |||
   |||
   |||
Db 2387 AAAGGAAAAGACCTCATCCAATGAATGTTCTCGAAGTGAATTTGGATGGGGATATCTAC 2446
   |||
   |||
   |||
QY 1014 TrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu 1033
   |||
   |||
   |||
Db 2447 TTTGTTGTTGGATCCAGATGATTCGCGCAAGGCAAGTCCAGTCGATGGATATATCT 2506
   |||
   |||
   |||
QY 1034 ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSer 1053
   |||
   |||
   |||
Db 2507 CCAGCACCCACACACAGTTGGATCATGAT----- 2536
   |||
   |||
   |||
QY 1054 HisGlyThrGlySerAlaAlaLysGluClnThrThrTyrAspMetIleGlnLysSerPhe 1073
   |||
   |||
   |||
Db 2537 -----GTCAATTTGAGGAAGTTGAAGAGTACTTC 2566
   |||
   |||
   |||
QY 1074 HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCys 1093
   |||
   |||
   |||
Db 2567 -----ACCAACTAT----- 2575
   |||
   |||
   |||
QY 1094 TyrIleAsnAsnSerValSer----- 1100
   |||
   |||
   |||
Db 2576 ATCATTAATGACAGTTTGGGAATCATGCAAAATCGCAGCTGTATTGGCAGACAGAGAA 2635
   |||
   |||
   |||
QY 1101 -----AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
   |||
   |||
   |||
Db 2636 CCTGATATGGCCATGAGGATCCGTGCAAACTTCTGCTCAGCTCTTTCAATTGCAAGTT 2695
   |||
   |||
   |||
QY 1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
   |||
   |||
   |||
Db 2696 GACTTTCCAAAGACTGGGCTTCCAGCTGAAATACCATCT-----CAGTTGGCGCTTAA 2749
   |||
   |||
   |||
QY 1136 LeuLeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeu 1154
   |||
   |||
   |||
Db 2750 GAATACCCAGACTTCATGAAAAAGCCAGACCAAGCAACCTATCCCTCGAAAAAGTTATT 2809
   |||
   |||
   |||
```

```
QY 1155 GlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAla 1174
   |||
   |||
   |||
Db 2810 GGAAG- 2815
   |||
   |||
   |||
QY 1175 IleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAsp 1194
   |||
   |||
   |||
Db 2816 -----CTTTTCCAGAAAGTGAAGACAAA----- 2839
   |||
   |||
   |||
QY 1195 GlyAlaHisPheTrpAspProAspLeuAlaSerTyrThrPhePheLysGluIleSer 1214
   |||
   |||
   |||
Db 2840 -----ACACCCAGGCTAGCTCCATCGCGACCTTCACAAAGGATGTTGCC 2884
   |||
   |||
   |||
QY 1215 AspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluVal 1234
   |||
   |||
   |||
Db 2885 AGGAATCATATGATAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2944
   |||
   |||
   |||
QY 1235 -----GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
   |||
   |||
   |||
Db 2945 TTTTACTACAAAAGTGAATAC-----GACCAACAGCTTGGTAATTTAATG 2989
   |||
   |||
   |||
QY 1252 AspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrProGlu 1271
   |||
   |||
   |||
Db 2990 GACTACTACGGCATAAACTCAG-----GCTGAATATCTTAGTGGT 3031
   |||
   |||
   |||
QY 1272 AlaMetAspLysSerGlyAlaAsnTyrAspSerLys----- 1283
   |||
   |||
   |||
Db 3032 GCATCATGAAAGCATCAAAAACCTTTTGACGCGCAAAAAGATGACAGGCCATTTGGTGT 3091
   |||
   |||
   |||
QY 1284 ValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp----- 1301
   |||
   |||
   |||
Db 3092 GCTGTGAGGTGTTTGAG-----AAGGAGCAAGGCCCTGGTTCAAG 3133
   |||
   |||
   |||
QY 1302 -----AlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyr 1314
   |||
   |||
   |||
Db 3134 AGCGGTAGTGATATTGATGACATGTTAGCAAGGCTTGGCTTGGTACCATGTTACATAT 3193
   |||
   |||
   |||
QY 1315 HisLysSer-----ProLysPhe 1320
   |||
   |||
   |||
Db 3194 CATCATATATATGCGGTCTCTACAATGAGGCTTGAAGAAGATCATTTCAATTAGCTTT 3253
   |||
   |||
   |||
QY 1321 ValTrpGlnMetAlaGlyArgGlnLeuLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
   |||
   |||
   |||
Db 3254 CCTGCTGTGTTTATCACCAGCTGATCCAGATTAGAGGCCAAAGCTCGTAAAGGCCG 3313
   |||
   |||
   |||

RESULT 5
AX204840 3591 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 2 from Patent WO0155407.
DEFINITION AX204840
ACCESSION AX204840
VERSION AX204840.1 GI:15394184
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Beclin C., Elmayer, T., Mourrain, P. and Vaucheret, H.
TITLE Novel 5982 plant gene and use thereof
JOURNAL Patent: WO 0155407-A 2 02-AUG-2001;
Rbio (FR) ; INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)
FEATURES
source
1. .3591
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
1. .3591
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60174.1"
/db_xref="GI:15394185"
/db_xref="EMBL:AC60174"
/translation="MGSGNKKSVTVQVSIIGFGESTTAKQLTDYLEVGVWRCR"

CDS
```

LQTSWTPGSPNPEIADTSNIPSIDVEYKVPFAVFAVESAGRAMDAAGCNLI
LDQGLKSLGKPNISNQRRTTVFYKLAGITLIEIGTLVSRDDFFVWRABGVDE
VDFPONTCKPFRKSTAFSDKADNVHAINCYDYLELLVRDIQTVRQIKLHGVFLI
QUASSERVYTDADDITDFTVPGDILLDDDDPWIRDTFQVGAIGRCHSVLISPRY
ENKLRGDFYFRMRVQBERVWPRIENRPFQFVSDHFFCIHKEGHSFEIMFLV
NSVLHGFQNPQLFERFDLIRNQKDVNTASLKHLCYKRPVDFDAYKRLKVQEWI
OKNPKLGSHEQSEDSISIRLVIPTRAYCLPPEVELSNRVRKYAVAEERFURVP
MDESQTLNSNLSVYVAIPVKDLTSSFSQKTVFVKVKSILTDGPKLGRKYSELA
FRANQJRBASAFRAEDKTRVSDIKTMWGFQKXNVAKCAARMGLGESYAYVDVM
PHEVTEPDIERNGVFSDGIGTITPDLEAVEVMEKLDVHVSFCAYQIYAGFKGV
VARWPSKSDGIRLALRDSMKFPFSKHTILEICSWTRFQFGLNQIITLUSVLGVDE
IFWMOQSLMYKLNRILODTSVAFELVASCABQNTAAIMLSAGFKPKEPHLGRML
SSVRITAGMLYKLEKRSIFVTSRWLMGLDRAEGLIHQOCFIOVSKPSIENCFKSHS
RPEKTYLDELVVGVYVAIPKPLHGVDRILEAVDVPOLHMHVDCILIFPOKGRPH
NEASGDLGDGLYFVADQKLIIPNRKSPYAPVYDAAEBSKLSGRAVNHODIIDFPAH
LANEQGLTICNVHVAHDSXGAMDESCILLAEATAVDPEPKTKIVSPERHLKPK
LTPDMFKGDIYQYKSNKILGRLYRVRVEVDEDEAASSESTDPISALPIDVALEIRG
FEDLPEAWGHKGLDGLIGLQYKVKQBEIIVTHGHSMPKITSKQGLBELK
HSYNLSKKEFRKVFETIPDHENLSEEEKNLYEKKASAWHYVTHPEWVKXLSLEOD
PDESSHAAMLSFAWIAADYLIARIKIRSREMSIDSAXFVDSLAKFLAQLR"

ORIGIN

Alignment Scores:
Pred. No.: 8, 0.2e-06 Length: 3591
Score: 328.00 Matches: 197
Percent Similarity: 37.17% Conservatives: 126
Best Local Similarity: 22.67% Mismatches: 312
Query Match: 4.43% Indels: 234
DB: 6 Gaps: 42

US-09-913-878A-2 (1-1402) x AX204840 (1-3591)

535 PhePheGluLeuLeuLeuProSerProThrSerThrSerProSerValProValVal 554
1396 TTTGTGCTCCTATTGCTGAAGGATTTGACATCAAGTTCTTCTCCAGAGACCTACGTT 1455
555 SerLysGluProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
1456 TTTAAAGAGTGAAGACATA-----TTAACCGATGG-----TTTAAA 1494
575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
1495 CTATGTGGTGAAGAATAC---AGTTTCTAGCATTTCTCAGCCAAT----- 1536
595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
1537 -----CAACTGAGA-----GACCGCTCTCATGGTTC 1563
615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
1564 TTTGCTGAACACGCG-----AAACACGT----- 1587
635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
1588 -----GTGTGATATAAAACATGGATG 1611
655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
1612 GGGAGTTTCAAGACAGAATG-----GCAAAATGTGCTCTAGGATGGGC 1659
674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
1660 CTGTGCTTCTCTCCACATATGCACTGTAGTGTCTATGCTCCTCAGCGTT----- 1710
694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
1711 GACACTGAGTTTCCAGATATTGAGANAATGGGTATGTTTCTCTGACCGAATGGTACA 1770
714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuAspVal----- 731
1771 ATCACACCTGACCTCGCTGACGAAGTAATGGAGAACTTAAGTTG---GATGTGCACTAC 1827
732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750

1828 AGCCCTTGTGCTTATCAGATACGTTACGCGAGGTTTCAAGAGGGTT----- 1872
751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln----- 765
1873 -----GTTGCTCGTGGCCATCAAAAAGTAGTGATGGAATCAGG 1908
766 -----ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
1909 CTAGCCCTTCGACACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGAGATC 1965
783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 802
1966 TGTTCCT---TGGACGAGGTTTCAACCTGGGTCTTAAATCGGCAGATAATTACCTCTTA 2022
803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsn 822
2023 TCC-----GTACTAGGTGTTCCGATGAAATATTCTGG 2055
823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
2056 GATATGCGAG-----GAATCCATGCTCTATAAACTGAACCGCATCTCTGTATGATACA 2106
843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
2107 GATGTGGCATTTGAAGTTCTACGGCATCATGTGCT----- 2142
863 PheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGly 882
2143 -----GAACAGGGAACACTGCAGCTATCATCTAGTCAGGT 2181
883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
2182 TTCAACCAAAACCCGACCGCATCTACGGGGATGTTGTTCTTCAGTCAGAAATGACACA 2241
903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
2242 CTCTGGGGTCTCAGAGAAAATCTCGTATTCTTTTACTTCAGAAAGGTGGTAAATGGGT 2301
923 IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer----- 940
2302 TGCTAGACAGACGCGGATCTGAACATGCGCAATGCTTTTATCAAGTCTCTAAACCG 2361
941 -----SerLysPheArgAspGluGluGluSerPhe 950
2362 TCTATAGAAATTTCTTCTCCAAACATGGGTCTCGTTTAAAGAGAGACAGATCTG 2421
951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
2422 GAAGTAGTTAAAGCTAT---GTAGCCATTGCTAAGAATCCCTTGTCTTACCCCGGGAT 2478
971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle 990
2479 GTAAGGATTTTAGAGCTGTTGATGTACCCACGTCATCATGTATGACTGCCTTATT 2538
991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyAspTyrAspGly 1010
2539 TTTCCCTCAGAAAGGTGATAGCGGCATACAAACAGAGCTTCTCGCATGTACCTTGACGG 2598
1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
2599 GACTGTACTTTTGGCTTGGGATCGGATCAGAACTCATC----- 2634
1031 ProLeuGluProAspLeuSerArgTyr-----LeuLysLysAspLysThrThrPheLys 1048
2635 -----CCTCCCAACAGAAAAGCTATCCGGCCATGATATGATGTCAGCTGAAGAAG 2688
1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMet 1068
2689 AGTTTA-----GGCCGTGCTGTCAACCCAGGACATAATCGATTTC 2730
1069 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsn--- 1087

Db 2731 TTGCGAAGAAC-----TTGGCGAATGACAGTGGGCACAAATTGGCAATGCA 2778
QY 1088 -----TyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProala 1104
Db 2779 CAGTCGTTTCATCTCATAGAGTGGAT-----GGAGCCATGGACGAGAAATGT 2829
QY 1105 IleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPhe 1124
Db 2830 TTGCTACTGGCAACTAGTCCTCATGAGTGTATTCCTCCAAAGACAGGGAATTTGTG 2889
QY 1125 AsnGluAlaSerTrpAlaGlnLeuArgGluLeuLeuGlyGlyAlaLeuSerLeuPro 1144
Db 2890 TCAATGCC-----TTCCACCTAAACCAACTCTCCAGATTTTCATGGGAAAGAA 2943
QY 1145 Asp---ProMetTyrLysSerAspSerTrpLeuGlyArg-----1156
Db 2944 GACTACCAACYTACAGTCGACAAATCTTGGTCTGCTTACAGACGGTAAAGAG 3003
QY 1157 -----IleAspTyrLeuLysPheSerIleAlaArgPro-----1173
Db 3004 GTTATGATGAGATGCAAGGCTTCTCAGAGAAGACACAGACCAAGTCCCATCCCT 3063
QY 1163 -----IleAspTyrLeuLysPheSerIleAlaArgPro-----1173
Db 3064 TATGACGCTGTTCTGAAATACCGGATTTGAGATTTGATCCCTGAGGCATGGGTAC 3123
QY 1174 -----AlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaLysAsp 1191
Db 3124 AAATGTTGTACACGGCAACTCATTTGCTCTTGGGCAATACAGGTGAGAAAGAG 3183
QY 1192 ThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhe 1210
Db 3184 GAAGAGATTGTACGGGTGATCTGCTGCTC-----ATGCCAAATACACAGCAAG 3234
QY 1211 LysGluIleSerAspLysSerArgSerSerAlaLeuPheThrThrLeuLysAsnArg 1230
Db 3235 AAACAAGGGCACTGAAGAAAGA-----CTGAAGCATCTT 3270
QY 1231 IleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSer 1250
Db 3271 TATAATTCCTTAAAGAGATC---AGGAAGTATTGTGAGAAACAACTCCCTGACCAT 3327
QY 1251 LysAspProTyrProValArgValAsnGlnValTyrGluLys-----TrpCys 1266
Db 3328 GAAATCTCAGCAGGAGGAGAAACATCTTGTATGAGAAAGAGCTTCAGCTTGGTAT 3387
QY 1267 AlaIleThr-----ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284
Db 3388 CATGTAACCTTACCATCCCGAGTGGGTGAAGAAGTCT-----3423
QY 1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304
Db 3424 -----TTAGAGCTGCAA-----GATCAGATGATGCTCAGTATCGCGGATG 3465
QY 1305 ArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMet 1324
Db 3466 CTGAGT-----TTTGTCTGG---ATT 3483
QY 1325 AlaGlyArgGlnLeuAlaTyrIleLys 1333
Db 3484 GCAGCTGATATTCTTGCAGAAATCAA 3510

RESULT 6
NCBI13B3/c 82584 bp DNA linear PLN 06-MAR-2003
LOCUS Neurospora crassa DNA linkage group II BAC contig B13B3.
DEFINITION BX284762
ACCESSION BX284762.1 GI:29881246
VERSION
KEYWORDS Neurospora crassa
SOURCE Neurospora crassa
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 Schulte, U., Aign, V., Hoheisel, J., Brandt, P., Partmann, B.,
Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 82584)
REFERENCE 2 (bases 1 to 82584)
AUTHORS German Neurospora genome project.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2003) MIPS, Institut fuer Bioinformatik,
GSF-Forschungszentrum fuer Umwelt- und Gesundheit, GmbH,
Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseleldorf,
E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT this contig is an assembly of BAC 13B3 from 1 to: 11950 and BAC
11C11 from 11951 to: 82584, strain OR74A, BAC clones are available
at the Fungal Genetic Stock Center http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
http://www.mwgdn.de

FEATURES
source 1..82584
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
/chromosome="6"
3959..3982
/note="24 bp tcatcatcatca tandem repeat"
5038..6880
/gene="B13B3.010"
join(5038..5145,5216..6880)
/gene="B13B3.010"
/note="similarity to clone mgc:8691, similar to amplified
in osteosarcoma, homo sapiens, TREMBL:BC000532.1"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD70506.1"
/db_xref="GI:29881247"
translating="MRPSLALSLSPFGSARQPASFSIHQDLAHPQFVIFSDS
YVFEADAFALLEANPKPTPASDGHGSTRTRDLTSAIREATANADTNGDESI
GTSPLREYELIAHPMYLCSIPITAPPALNTATLAKAEAREVTRYNKGWE
LMGLENQCLFVSGWMSYQYCYKSIQVHVAVNPKNPGLDKNSQVILGTS.LPP
SSHSQKGOKEIVENNEKQLSPPTNLEQAKNORYLVORLDGTCIDLTGRPTTET
QYHCNALSQDRICNKEVTTTCAYLVITPRLCAOVAELPPKXKAPHTCTCHITIS
DEBALFSQRKRTIDSAAAAAAATTTEDQKQSGESGKLSYQSITVAGIFIGARR
ILPSTHVLPLPHLQOQROAQQNLLEALTAAFKADVDFGDDNNNNNNHHHPKA
GKRGKAGAGKGGQGGKMKRISERIDKLGDDQTLDALEETRAAGLDPPRLN
DERAGGEIWFYADVSGDEDDGVAEREGKEFVFWYDEDEGGEPAEAGKQKESKK
GGNGSGSGSBEBSKEEYVRDEL"
5038..5145
/gene="B13B3.010"
number=1
5146..5215
/gene="B13B3.010"
number=1
5216..6880
/gene="B13B3.010"
number=2
7534..7562
/note="29 bp poly(a) tandem repeat"
12653..15824
/gene="B13B3.020"
join(12653..12719,12910..13083,13261..13438,13497..14132,
14195..15824)
/gene="B13B3.020"
/note="similarity to transcription activator, Aspergillus
oryzae, PIR:JC7252
[VCDKRLRRIRDYQSPQCIDAKLQC]"
/codon_start=1
/product="related to transcription activator"
/protein_id="CAD70507.1"


```
/db_xref="GI:28881248"
/translation="MMFMWGAHRRRRNLAILPGLQRLGSSPEAQIVRLKTFPGTS
TKQSPRSSNATAGSNNRPVQPKVCKRLRIRCDYQSPQCIDAKLQCTR
NHVPRKRPGRGLINLRERAEKKEQEPSSGDPVSDGARNNTDIYPQNAS
VSTWASDVSTAPETGLWANEHIIPSRGAFYILPRCVLFYQHLVPMVMPPI
QMDRVNCLPEKNLFLSLATCFRMSGHLSGAEDSPFWQAGRFILNCLDVRKO
YLDYDNISLSTVSSIFLASSFETNQSKAWVLREALTFAQLGLEDESTYAGLSP
EALCQRVFWLLIWNERSFALLRNKPLMLRAPKLPMTLHAYESRGIHTGFLQLGI
YPLRDSIIETWYSGNSTVDNTYLALONLARPGSSGGLSIPHPFPSPNAVS
SIQMPVSCSDFSLSSSSSSPDEVPVVEFTASQTAELLVTQWRLIILWS
SIHQGLYHAAENSMHFALPLTIARQTLGLRLSLMQIGGLEATGVAALVNGMGIFE
KIFETGAMCNVLDYKASLEGQTLPLNFRHRHSAASGNKQARASATSSPADQM
YLDGEDVLLDFVRLASATPSKQFAEPLMYWTRPGMRIGPGSVLHLDTGCG
TGAFMATQPLASQIPLGLNGLGLDHAAGLCAAGHDASSNMHQOQPGIWNAA
SAYAGQAPLPSATGSAUGLSPRPGPGPGGQGGMDYVNGQASHQGGGRNIDR
LGAGSNVRLGLOHATFGARPLRNAPLPLPFGPGGPTSLYDVPNSFIDPFGGLT
RARSRTMGQGLRGJLAMYVQGGPGTGGSGSPSVSQAS"
12653..12719
/gene="B13B3.020"
/number=1
12720..12909
/gene="B13B3.020"
/number=1
12910..13083
/gene="B13B3.020"
/number=2
13084..13260
/gene="B13B3.020"
/number=2
13261..13438
/gene="B13B3.020"
/number=3
13439..13496
/gene="B13B3.020"
/number=3
13497..14132
/gene="B13B3.020"
/number=4
14133..14194
/gene="B13B3.020"
/number=4
14195..15824
/gene="B13B3.020"
/number=5
16797..17552
/gene="B13B3.030"
complement(join(16797..16985,17226..17552))
/gene="B13B3.030"
complement(join(16797..16985,17226..17552))
/gene="B13B3.030"
/codon_start=1
/product="hypothetical protein"
/db_xref="GI:28881249"
/translation="WHFKSLLSGLLSLLRYHAALPTISDDTINDDTTINDDTLI
PRVMPKDPNNMFDNNKFNVTYSITONPDHDTLICPSRALLSGPPPEPGRFST
TLGRWTVSWTQNAPTYKTLAEQAWLDQEKRTTATNTRELPLVAGEAYVGHDAW
SVVLKGVVDC"
complement(16797..16985)
/gene="B13B3.030"
/number=1
complement(16986..17225)
/gene="B13B3.030"
/number=1
17029..17053
/notes="25 bp poly(t) tandem repeat"
complement(17226..17552)
/gene="B13B3.030"
/number=2
18460..24617
/gene="B13B3.040"
complement(join(18460..24111,24165..24617))
/gene="B13B3.040"
```

CDS

```
complement(join(18460..24111,24165..24617))
/gene="B13B3.040"
/notes="similarity to am-toxin synthetase (amt), alternaria
alternata, TREMBL:AF184074.1
Contains Putative AMP-binding domain signature
[IQFTSGTGKPK]/ATP/GTP-binding site motif A (p-loop)
[AWSGSKT]; Cysteine proteases inhibitors signature
[QQVQCALSFGLN]; Phosphopantetheine attachment site
[AVGGDSLRAKMLRLAAA]"
/codon_start=1
/product="related to AM-toxin synthetase (AMT)"
/protein_id="CAD70509.1"
/db_xref="GI:28881250"
/translation="MOQVVGQALSFGLNGQSSNATSHLDTLGLTFSRPLSPQK
NLSLDSALLGILLLOREGRDDRVENFMGRRTSDGAETSSQFSKVALSGELDQ
TASVKNVLVLKVTQDVTENRILPNDGKEAPEKIEQEVYEAQRPNWKFOLELL
ASDGGHDLITQAFWKNVPDALTSSQVITRLNATFELLGLIFSPETTPSSLLAP
LPCDLNQWAFNAWVPLDITRMQDIISQAANPSKPAVSWDGEFTFAEGRKSDL
LARHLVSRGVKLGVVPLMFERSMTIIVALLAMKAGSAFALDPTFSQVEKRVNVE
QTGGDIIVASQTELAELVPLSSSGSQGVITVSDDLFTLSOTPELTSVTLVPV
TOTPLYIQFTSGTGKPGVISHANTYSGALPRAEAVGYKSSRCFFSFAGFVDS
IDCMLCTAVGGTICIPSDADRMNDLGGAILKSGANMAHMTPSVARVLDEVIKGLDV
LGLGSAASAADAANKSKTSVIYAGSECTVGCTVNNVFARDEAKKGGFTGNI
GTCVGGCGWVPDNDHRLVPVGAAGELLMEGVGLGYLEEETKEVFIEDPKHLV
AGHDAPGRKGRLYKTDLYDADGSAFVVGKQKQAVKLGRVSLVETIEHLR
AKLPBGKIAAEVIRKPSGGDPTLVAETSTGISAKDALQSVVEGVSFSELQAA
LDGIERVLGIDLPYMPNAYIFLRMPSLVSGKIDRKTLRAMGAMTRERQVSGSAX
ARQGGEGAPTTEMETALHKWKQLIGTQEAAGDSFVAGDSLRAMELIAAARAE
GVKLTADVFRVPLDMAVATKVDQDAGKDNNEIAPFSLLEDMAIEDVKAFAEAS
KXNLDKDALEDVYPCFLOEALMALSAAKENVYVAQVRLKMGQAADKLAAPAI
NADCALRTRIHSRGLMQVVVKEPLHWSATTLAILETRNEEMDLGKLVRYGL
IQDGDYHIVLTWHALYDGMPLVDRVNOAYQGLVPRKPAAFKFKFINTLNITLD
RAADAYKWDLAGATGVQFPALPFGYQTRADSLWELVSLDGRKPPSPNTITLA
TIVRAALVAEQYTAGNNDIVFGETLTGRNAPIVGADEIEGMITTVPRVICIDREQ
TVAEYLOQVAEQMITQIYFHAGLQHIRLSDDIALQACELRTGIVLHPAAGEVGTEN
TPANGIAPAGDSEAQAELFNNTYALMLVCSLEADGLVMAFSFKTBMVDVVVKALD
QLOVVVQLCEDGCGEKKVQLECLTEKDHEVSKVLGAVKNTLAESDPVGVGTAEVD
EGWIVDAADTARLVVCAAGELIVETSELDAPVTLLAETPAWTEGKCLKYTGOLA
KFKYDGSASLILKASEPTEVAKKVAASVTPSPAATSAKQTLIRKIKWALKVKV
EDSIYLGDSFFNRGDSITAMKLVSERQOGLKISVAQVFNARTLFDMANVQTTGTV
VDASEKVEYKFPFSLLENNKASINEVVVHSLADSKTIADILPRIQEIAVRGTI
ELPFRSLRYELHFNQVQDKAALFRACHEVARNELRTVFVRDASNTCYGAVINDLI
TSTAIEYIDHDADVDFAAQVCRLDSTQRMYPGVSFVFWFFITNGTSSQSLAPRLSHA
QYDEICLPIMLHQHOLYNPESSTPPSPYRFSHFVAHTLRDSIPAIIPYMNLLSGS
SGITRPRPTPTITSRKHFAIHRADVLISSRPDVTATLPSAAWALALASHVSPDLF
```

Alignment Scores:

| | | | |
|------------------------|---------|---------------|-------|
| Pred. No.: | 0.00462 | Length: | 82584 |
| Score: | 306.00 | Matches: | 274 |
| Percent Similarity: | 33.15% | Conservative: | 143 |
| Best Local Similarity: | 21.78% | Mismatches: | 440 |
| Query Match: | 4.14% | Indels: | 402 |
| DB: | 8 | Gaps: | 57 |

US-09-913-878A-2 (1-1402) x NCB13B3 (1-82584)

| | | | |
|----|-------|--|-------|
| Qy | 132 | ProProProAsnAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGly | 151 |
| Db | 45061 | CTCATCCCAACTCTCAAGGAGAACCAGCCGACAGCTTTCAGGACGCCCGG | 45002 |
| Qy | 152 | LeuSerArgPro | 156 |
| Db | 45001 | CTGCGTTCCCGGTGTCACCTCCCGCTGCTGCTCTCTCTGTTGTTGTTCCAGGAC | 44942 |
| Qy | 157 | SerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspPro | 174 |
| Db | 44941 | CACGAGCTTCAACGGCCCTGCTCGTCTGTTCTCGTGTGTTTTCACACACACTGACG | 44882 |
| Qy | 175 | -----ProLysArgSerLeuThrArgSerAlaThrGlyPro | 186 |
| Db | 44881 | GGTTCAACCCCTCTGTTGTTCCGTCGCCGCCACCAAGCTTTCAGGTCCTCAATTTGCTC | 44822 |
| Qy | 187 | ProIleHisGlyAlaAlaIleProLeuLysPheProAspProValAsnThrGlySerLys | 206 |

| | | | | | | | | | | |
|----|-------|--|--|--|--|-------|----------------|---------|---------------|-------|
| Db | 44821 | AAGTTGAG----- | | | | CTTTT | GACACTCACTTCAT | TGGTCAA | CCTTTGGTGAGCT | 44774 |
| Qy | 207 | ArgProSerLeuGluSerGluAsnLeuAenGlnCys- | | | | | | | | 218 |
| Db | 44773 | TCTCCCTTCTCTCAAACATGTGTTGATCCAAATAATGTGCACACCACATAGCATCCTAAA | | | | | | | | 44714 |
| Qy | 219 | -----ThrLysBargalAlaLysGlyLysLeuSerASP- | | | | | | | | 228 |
| Db | 44713 | TAGCTGGTCGACAGCTCTATCGCCAAATTATCCTCTCGCCGCCCATGGAGGTCTACATCA | | | | | | | | 44654 |
| Qy | 228 | ----- | | | | | | | | 228 |
| Db | 44653 | GGRACCTTCTCTCAAGGCCTTTCGACAGAGAAGCTTGACAGCTCATCTGAAGCGGTCATGG | | | | | | | | 44594 |
| Qy | 229 | -----AsnValAlaAlaAlaAlaProProValProIleAlaSer | | | | | | | | 242 |
| Db | 44593 | ACAACCTTGGCATCCGCTCACTAGCACTGCAATAAAGCGTCCAACCAACCCCTTTGGGAAG | | | | | | | | 44534 |
| Qy | 243 | AlaLeuAspLysValProThrArgArgHisAlaAsnThrArgASP-ProThrAlaThrGl | | | | | | | | 262 |
| Db | 44533 | TCATCTTTATCATCTGCCAGGAGCGCTGAACCTTCATCGCAAGCACCCCTCGTCTTTATA | | | | | | | | 44474 |
| Qy | 262 | yHIS-ArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyrgLyse | | | | | | | | 282 |
| Db | 44473 | TCATGAATCAGCAGGCTGAATGCAGGT-----CAGCAAGCCCCAGCAAGGATC | | | | | | | | 44423 |
| Qy | 282 | rSerValPheSerAlaCyArghisAngInSerThrGlnSerSer-PheGluAlap | | | | | | | | 302 |
| Db | 44422 | TGAAGAACTTGGTCAA---CGGCAT-----TCCTTACGAGGCTG | | | | | | | | 44387 |
| Qy | 302 | roProSerGlnProArgGluLys-----ArgPro- | | | | | | | | 311 |
| Db | 44386 | AAGAAAGGAAGCGCAAGAGCGCTGGAGGAGAATACGAGCAACAATGCCCGCCTG | | | | | | | | 44327 |
| Qy | 312 | -----ValAspAlaThrValPheGluAlaGlyHis | | | | | | | | 321 |
| Db | 44326 | CAGCGACGCCAATTTAGCACTTGATCGCGAGGTTCTTGAC'TGCGGACATTTGGTCTCTACA | | | | | | | | 44267 |
| Qy | 322 | -----LeuileGluSerProSerLysGlyArgThrThryLysSerHisI | | | | | | | | 336 |
| Db | 44266 | TCGACGGCACTTGACCTTCATCACCAGTAGTGTCTCTCCCTGGCGCTCTCAGCAAGT | | | | | | | | 44207 |
| Qy | 336 | leAspAengInProLeuSerSerSerGlnGlyGluThrSerPheSerThrytyr- | | | | | | | | 355 |
| Db | 44206 | TCGCCAAGCAGCACCTCGTCATTACCTCTTTGGCTGAATCTCAGGTGTCTATCCCATATC | | | | | | | | 44147 |
| Qy | 356 | -----GluSerPheProSerSerGlyGlyGluAlaIleProGluProSerA | | | | | | | | 372 |
| Db | 44146 | GATATATCCACGACCTGTTTGGTTCGAATGGTGACATGTCCTGTTC----- | | | | | | | | 44100 |
| Qy | 372 | rgSerAengGlyLeuAlaArgSerGluGluSerAlaArgSerGlnValGlnValHisAlap | | | | | | | | 392 |
| Db | 44099 | -----ACTCTGCGACACGCCC | | | | | | | | 44084 |
| Qy | 392 | roVal----ValAlaAlaArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHis- | | | | | | | | 410 |
| Db | 44083 | CTAGTTTCTCTCGCCGCCAATTCCATTACTTGGCCCGGCAACCCCAAGAGGATGCGCC | | | | | | | | 44024 |
| Qy | 411 | --GluAlaProLeuAlaValAlaIleTrpGluValThrArg-----LeuPheMetHisCysL | | | | | | | | 428 |
| Db | 44023 | TCGAGGCATGTGATAGTAACACGCCCAAGATATCAAAATTCGCTCTCTCATTTCA | | | | | | | | 43964 |
| Qy | 428 | ysValAspLeuGluAspGluSerLeu-----GlyLeuLysT | | | | | | | | 440 |
| Db | 43963 | AGGTT-----TCGGCAACAACATTAAAAACAGAACCCAGGCTAGCTACAGAGGGAGTGAT | | | | | | | | 43910 |
| Qy | 440 | yrAspProSerTrpSerThrAlaArgAspValThrAspiletTrpLysThrLeuTyArgL | | | | | | | | 460 |
| Db | 43909 | TCCGTACCGGCATTATGGAAGTGAATGACGAGAAATGTTCTGGGTACGCATTTATACAT | | | | | | | | 43850 |
| Qy | 460 | euaSpAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThra | | | | | | | | 480 |

| | | |
|-------|--|-------|
| 43849 | TC-----GCCATTCAACACACATCTCCGGAAGATGCATATGCAGTTG | 43800 |
| 480 | laMetThrGlyAsnPheGlySerLysGlySerAlaValValLeuSerAlaValLeuAspTrp | 500 |
| 43807 | CAGTTGACAGG-----CTTAGGGCGGCTCTCTCTGGAGT | 43775 |
| 500 | yrAsnProAspAsnSer----- | 505 |
| 43774 | ACGAAGGTAGAGATACGTTCCCTTACAGCGTGCCTTTCAATCTCAACGCGTGGTGTACA | 43715 |
| 506 | -----ProThrAlaProLeuTyrLeuValValLysLeuLysProLeuMetPheG | 521 |
| 43714 | TGTCGTATCTTCATCCGACACACCGTCTAAAGCTAGCCACGCGCTTAGCTGATGTTCTG | 43655 |
| 521 | luGlnGlyCysArgLeuThrArgArgPheGly-ProAspArgPhePheGluLeuLeuLe | 540 |
| 43654 | AAACGCCCGACGCTCAGGTCCAGCGCAGGATCCCAATTTCGG-----TTGATG | 43607 |
| 541 | ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyVala | 560 |
| 43606 | CTTTCAGGACCTCTTCAAAACCACTGACTGG-CCG-----TCGCCGAGCCTCTCGAGC | 43554 |
| 561 | ValGluGluValIleGlnTrpLeuThrMetGly-----GlnHis-----SerLeuValGly | 577 |
| 43553 | GCAGAGAGCTCCTCAGTTTTCAGGTGCGAAGGATATTGAACATCTGAAGAAACGGAG | 43494 |
| 578 | ArgGlnTrpArgAlaPhePheAlaLysAspAlaGly----- | 589 |
| 43493 | AAACGATCGGTGAGGGCTACACCTCGGTTAAAGAAAGAGATTCTCCCGTCTTACC | 43434 |
| 590 | -----TyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysPheProIle | 607 |
| 43433 | AAGATCTATCGGCACTTGTCTACACCAACCCGAATCGAACTTCACGGGCTCGAACTGGAA | 43374 |
| 608 | IleLysGluArgVal-----HisPhe----- | 614 |
| 43373 | GCCAAGAACAGGATCCTTCGAAAGTTTCAGAACACCGAGCACCTTCTTCGAGTCCAG | 43314 |
| 615 | PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro | 634 |
| 43313 | TTCCGCCAGAGGAT-----GGACAGGACCTCTTTTCAACTCGGCGGT----- | 43269 |
| 635 | AlaGluGluProValGluGlnArgThrGluPheLys----- | 646 |
| 43268 | TCCATGGACGCCATATATACAGAGA-----TTTAAAGACGCTTTGAACCATGGCATCTCA | 43215 |
| 647 | -----ValSerGlnMetLeuAsp----- | 652 |
| 43214 | GTTGGTGGACGGTTTAGAGTTTCTCGCTTCTCCCATTCCTCCCTCCGAGCACATAGC | 43155 |
| 653 | ----TriLeu-----LeuGlnLeuAspAsnAsnThrTrp | 662 |
| 43154 | CTTTGGCTACGCGCGCTTTCATATACACGCGTAAATTGCAATTAGCTAGTAAACATCATT | 43095 |
| 663 | Gln-----ProHisLeuLysLeuPheSerArgIleGln | 673 |
| 43094 | GAGATTGGGCGATTTTCGCAATATCATGTGCCGACGCGCACGTGGCGCAGCATTAGGC | 43035 |
| 674 | LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis | 693 |
| 43034 | CAGGCTTTTCAGAGACACCATATTCGGTTTCGCTCTACGATCAACGGTATT----- | 42984 |
| 694 | LysThrAspLeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGly | 710 |
| 42983 | -----GATGTGATAGGCAAAAGGACGTCMAACCCAACGAAACGGGTGTTCCAGTGATGGC | 42930 |
| 711 | ValGlyArgMetSerArgSerValAlaLysArgIle-----ArgAspValLeuGlyLeuGly | 729 |
| 42929 | GTTGGGATAATTTCGAGGGTCCCTCGAGGTGATTTCATCGCGAGATTCCCGAATCCCAAG | 42870 |
| 730 | AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp | 749 |
| 42869 | GGCTATCCCAATTGCTACAGTTTCATGGCTCGAGCAAAAGGCGATGTTGGCTTCGAT | 42810 |

```

QY 750 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 769
Db 42809 GCCCGCTCAGTGGGAGCAATCTGTATCAGG-----GATTCATCGAGAAG----- 42762
QY 770 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 789
Db 42761 -----TTCCGAGCGCGTATCGAGAACATCTCGAAATATCGCATCGCTTCCAAAGCCA 42708
QY 790 LysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLys 809
Db 42707 ATTCCTCTGATGCTGAACCGTCAGATGATCAAAATCTTCGAGGAC----- 42663
QY 810 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgLysPheSer 829
Db 42663 ----- 42663
QY 830 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 849
Db 42662 -----ATGCGCGCCCTGCC-----CAGTGGTCTTGGAGCTTCAA 42627
QY 850 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 869
Db 42626 GAAAGGAACCTCAACGT-----CTCGCGGCAATCACGGAT 42591
QY 870 SerGlnGluGluThrLeuAsnPheLeu----- 878
Db 42590 AACGTCGAACGTGCAACCTTTCTCAAGCTGCGAGTGTGCGAGATCTGTACACCTC 42531
QY 879 -----MetAsnSerGlyPheAspProLysGlnLys 889
Db 42530 AGCCAGTTTCTCAAGACCTTGACAAATGAAC-----ATCGACTACCGGAGACACG 42477
QY 890 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 909
Db 42476 TTTCTTCGCGGTATTCTCGAAGCTGCTGCTTAGAGCTCGCTGCTTCAAGACACAG 42417
QY 910 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 929
Db 42416 GCTAGAATCCGGTTCCTTATGCGGTACTCTGTGTGTTGTCATGAGACGACAGCCTA 42357
QY 930 LeuGluGlnAsnGluValHisValGlyPheSerSer-----LysPheArgAspGlu 946
Db 42356 TTCCGCGAAGCGCAAGTATACGTAATTTGACCGTGGAGCGGTGCTTCAAGACCCCT 42297
QY 947 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 966
Db 42296 CCCACTGCT-----GGCCCCGTTGTGTGTCACACAGGTCTCCAGCCTTG 42255
QY 967 PheProSerAspIleGlnArgValArgAlaValPheLysPro-----GluLeuHis 983
Db 42254 CACCCTGGAGACATCCAGATTGCCCAATGCCAATGCCCGGAGGACCCCTCTCAGA 42195
QY 984 SerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAlaLysLysLeu 1003
Db 42194 GAACCTGAAACCTGCATCGTGTTCAGTCAAAACGGTGAACGTCACCTGCTAGTCAGCTG 42135
QY 1004 SerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal--- 1022
Db 42134 TCGGCGCGTGACTTGGATGCGGACACTTTCAACGTGATCTGGATCAAAAGCATCGTGGC 42075
QY 1023 -----AspGlyPheValAsnAlaGluMetPro---LeuGluPro---AspLeuSerArg 1038
Db 42074 ACGCTCAGAACTTTCGGCGCGCGCGATATCCACGTGTCGAGCCCTTGAACATGAACCGC 42015
QY 1039 TyrLeuLys---LysAspLysThrThrPhe-----LysGlnLeuMetAlaSerHisGly 1055
Db 42014 GAGTTCAGTCAAGGAGCATGGTGTACTTCTCGTGAGTTTATGAAGCCGATCACCTG 41955
QY 1056 ThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPhe 1075
Db 41954 GGGTAATTGCGGTTCGGACATGATCTCCGCGAC----- 41919

```

```

QY 1076 AlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIle 1095
Db 41918 -----GAACGC----- 41913
QY 1096 AsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db 41912 AATGAGGGAACCTGGATCGACACTGTCTCAGCTGGCAGCATTGCACCTAAAGCGGT 41853
QY 1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
Db 41852 GACTTCTCAAAATCGGGGATT-----CAGCTGGATATCACCGAGTTACCAAGG--- 41805
QY 1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp 1151
Db 41804 -----CCACCAATGATCGGCTGAT 41784

RESULT 7
LOCUS AY029284 6847 bp DNA linear PLN 28-DEC-2001
DEFINITION Neurospora crassa suppressor of ascus dominance (sad-1) gene,
complete cds.
ACCESSION AY029284
VERSION AY029284.2 GI:13699900
KEYWORDS sad-1
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 6847)
AUTHORS Shiu,P.K., Raju,N.B., Zickler,D. and Metznerberg,R.L.
TITLE Meiotic Silencing by Unpaired DNA
JOURNAL Cell 107 (7), 905-916 (2001)
PUBMED 11779466
REFERENCE 2 (bases 1 to 6847)
AUTHORS Shiu,P.K.T., Raju,N.B. and Metznerberg,R.L.
TITLE Direct Submission
JOURNAL Submitted (06-APR-2001) Biological Sciences, Stanford University,
385 Serra Mall, Stanford, CA 94305-5020, USA
REFERENCE 3 (bases 1 to 6847)
AUTHORS Shiu,P.K.T., Raju,N.B. and Metznerberg,R.L.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Biological Sciences, Stanford University,
385 Serra Mall, Stanford, CA 94305-5020, USA
REMARK Sequence update by submitter
COMMENT On Apr 20, 2001 this sequence version replaced gi:13641450.
FEATURES
location:Qualifiers
source
1. 6847
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
/chromosome="I"
/map="between leu-3 and cyt-1"
997..>6375
/gene="sad-1"
CAAT_signal
997..1001
/gene="sad-1"
TATA_signal
1119..1125
/gene="sad-1"
mRNA
join(1204..4542,4601..>6375)
/gene="sad-1"
/product="suppressor of ascus dominance"
join(1401..4542,4601..6375)
/gene="sad-1"
/note="SAD-1; RNA-dependent RNA polymerase; required for
meiotic gene silencing of unpaired DNA"
/codon_start=1
/product="suppressor of ascus dominance"
/protein_id="AAK31733.1"
/db_xref="GI:13641451"
/translation="MSGIPNWVQAQHAIIIVSDRSPKQDAPRPHRRSSQOQHR
ENTTPRPPSRPSPPSSPPKPPFAPQFSGKPSAPSLPSLTQVPFISREWKA
RTAIRVKDVPRAKLQDLHDLFSEYGHISYIELDEDRQVSDRGRRALIRFPFPLPS
TDFLCRCIGRLIGLVHMLPLETVDPKREDRKERTIKTGLNCTPQTLVFPSTLSF

```

GLLVQPAVFMKQSQVQTSSDSFLERDFDKMRVLVIFKSLHYQRELSGFRQRNHYKX
 LHKFGVTKELCTVGVSEHQAQVITLTDPPVAYKRDVSKTFFQEDRLTWSNDLWE
 RYVDSPLGDSKFNVAEHNQYIDLGLWLTWIELDQQTVMVDQVQVQLLDNWL
 TKLVTFPEPLNPKQKWDVFLDDRYKHDIQVSSRSNSNDPSLLAAPPRIISLPFDVRY
 QLEVCISQIINEHNIIDPFLEKMEFQNDONSFGKDRARLILEVADEYAGKIFDP
 MELPKDHALAVPEPTSMIPNHCAMVRVTITPTRIYFSTPCVETNRVIRQWKHAOD
 YFIRLOFDEVELEGKSGEAEPLPLFVAVVLEKGVAMGPHMKFLPAGNSQIREAG
 AMFCEQSNLTGDMRMAMGRFHHIKVIAKTAARLUGQCFTIRLVGPAPRIVITPD
 VEKDFCFDVGKISPLLAKIVADHWSIDPPFSAYQFMGGCKGVIVTVPDVKMEV
 HIRKQEFVAFENGLEVVRCQSFSTALNRIQIAVLSLGVDPQVQFVDMMEQQLSDP
 NAAMEDKQKATAIKTFIDENHMTPIIAEMLAYFMGQSEPFVFTLLQLMRSNFKTL
 KEKARLNVEKSAFVLGCVDETGTLGHMKVIEDKWDVSEKLPQIFLQIPDDVANGVYR
 VTITGVGRNPLSHPDRIYVVEADVDPALRHLDVVFPLTGDRDVSMSGGGLDQ
 DDFVWMDPLIPKERSHPMLSEPIACKELATEPTVANLITFFVLVYKXNLPLIHH
 AHLATADAEEVGVKPKCLELASLASHVADVYKTVGAEPFRRLDPKTPWPFNKH
 THSVTALGKIDYMKRETDFNMKENTQFPNRIUKTKCALRDGTILAKARRIKSQY
 DTAMRVQCLBIATIEFEVFTAVFMSKPRVGSDDYKLDNVGRESALKQHPKDKCKE
 AGDLLFSVAMRYTYEVRALFEAKQPHVPRDGRIGTRKTKTPKMLPLVSPFWLFW
 DKLGELARAGVLORLDQSDMDLLSDVPLVRSORRKHNSGSSDFMDEHGDPLS
 YTRSDGKTIHVGQILNLPFSDHDDGDERNARDNSSEDSYHVSNSKSSNLSPV
 AEEDLLTESPPASGVTPASPOVDLLGPTLTAKELAIAGLGPVATYTPPTDEE
 TVICGHENRASHLSASSLDPOPLIBETSENEEDLLDIYSASPPRADARLASAV
 GSDVPVKGFPFVPMQVIRMGHRIPTPLDSIAIPNVIGVFPESPLGTFTVASALIQ
 DPFVSPSPAVATPATSGSGGWMGGGGGGGGYIGIGGPFINKELVMDLGGIKBEDRE
 ETDDEAEVELEBIDEEDPIAVRYAQMAAL"

ORIGIN

Alignment Scores:
 Pred. No.: 0.000373 Length: 6847
 Score: 299.00 Matches: 302
 Percent Similarity: 33.44% Conservative: 206
 Best Local Similarity: 19.88% Mismatches: 510
 Query Match: 4.04% Indels: 506
 DB: 8 Gaps: 67

US-09-913-878a-2 (1-1402) x AY029284 (1-6847)

Qy 89 AsnTrpValProLysAla---HisAla-----AspProAspThrLeuProTrpSer 104
 Db 1416 AACTGGGTAGACGAGCGAGCATCCATCTTGTGTGCGCGGACAGAACGCCAAGAT 1475
 Qy 105 LysGluPro-----ProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGln 120
 Db 1476 CAGGAGCGCTGCGGAGGATCCTAGCGCGCCAGATCCAGCCAGCAGGACATCGAGAG 1535
 Qy 121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
 Db 1536 -----AACACCCGACACCTCGACACCCTCGGACCA----- 1568
 Qy 141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
 Db 1569 -----ACCTCGGACCCCTCCTCCAGCTCTCCC 1595
 Qy 161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
 Db 1596 AAG-----CCTCCCTCGCGCTTTCCG 1619
 Qy 181 ArgSerAlaThrGlyProProLeuHisGlyAlaAlaLeuProLeuLysPheProAspPro 200
 Db 1620 CAGTTCTCAGCGGGAACACCGCCTCATTCGATCCCTCC-----TCTCTGACACACCA 1676
 Qy 201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
 Db 1677 GTCTTCATTCGAGGAGTGG-----AAGTGCCTCAGG 1709
 Qy 221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaProProValPro--- 239
 Db 1710 ACCGCTGCCATTCGGCTCAAGGAT---GTGCGCGCTAGAGCTAA-ACCTCCAGGACCTTCA 1765
 Qy 239 ----- 239
 Db 1766 CGATCTCTCTGTGAGTATGGTCATATTAGCTACATTTGAATTGATGAGGACCGTCAAGA 1825

Qy 240 -----IleAlaSerAlaLeuAspLys 246
 Db 1826 TGTTTCTGATCGTGGCGGAGAGCTCTGATTCGCTTTGAGCCACCTCCGCTATCCGA 1885
 Qy 247 ValProThrArgArgHisAlaAsnThrArgAsp----- 257
 Db 1886 CTTTCTCTGCGGAGGATCTGCAGACTCAGAGTTGGACTTGTGGACCACCTGGCTGCTCT 1945
 Qy 258 -----ProThrAlaThrGlyHisArgAlaAspGlnValAsp----- 270
 Db 1946 CGAGACAGTGGACCCCAAGCGGAGGATCGCAAGAACGTCACCATCAAGACTCAGCTCGG 2005
 Qy 271 -----SerPheAspThrSerGlnGlyThrSerGlyThrSerGlySerVal 284
 Db 2006 CAACACTTGGCCGCAAACTCTATAGCTCTTCCCTCAACACTCTCTTT---CGGTCTCTT 2062
 Qy 285 PheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSer 304
 Db 2063 AGTCAGCCTCGGTATTCATGAAGAAACAATCCGTTCAAA-----CCCTCT 2109
 Qy 305 GlnProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuLeuGlu 324
 Db 2110 CGTCGAGCTCTTTCCTGAGACTTGGTTCGATTTCAAGCGCATCGGACTCGTGAATAAAT 2169
 Qy 325 SerProSerLysGlyArgThrThrLysSerHisIle-----AspAsnGlnPro 340
 Db 2170 TCTCCCTCC-----ACTACCAACGCACTTAGTGGGTTTCGACAGCAAAATGAGC 2220
 Qy 341 LeuSerSerSerSerGlnGlyLysThrSerPheSerThrTyThrGluSerPheProSer 360
 Db 2221 ATTCAAGCTCCACATTAAGTTGGAGTTCATCAAGAGCTTATCAGACCAACCATCGG----- 2274
 Qy 361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
 Db 2274 ----- 2274
 Qy 381 GluSerAlaArgSerGlnValGlnValHisAla-----ProValValAla 395
 Db 2275 ---TCGCGGAGGAGCATCGTCAAGCTTTG-GTCATAACTCTCGGACCCCGCGGTGCG 2330
 Qy 396 AlaArgLeuArgAsnIle-----TrpProLysPhe 405
 Db 2331 TACAGGAAGAGGATGCTCAAGACATTTGGCGAGATCGGCTGACTGGAGGAGAT 2390
 Qy 406 ProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArg----- 422
 Db 2391 GACCTCTGGGAGAGAGTGGTCGACATTAGCCCTGGAGCTCGATGTACGCAAAAACCCAGTT 2450
 Qy 423 -----LeuPheMethIscysLysValAsp 430
 Db 2451 TCCTGCGCGAGAACCCCAATATATTGACCTGGAGGCTGGCTTACCTATTGGATTGAA 2510
 Qy 431 LeuGluAspGluSerLeuGlyLeuLysTyArgProSerTrpSerThrAlaArgAspVal 450
 Db 2511 CTGGATCAACAATCAACAAGAGTC-----TGGGATCAAGTTCAACAATAT 2555
 Qy 451 ThrAspIleTrpLysThrLeuTyArgLeuAspAlaPheArgGlyLysProPhePro--- 469
 Db 2556 TTGCTCGATTGGAATCTCAGAGCTAAGTTGACTGTGTTTC---CCCAGCGCTTCCCAAC 2612
 Qy 470 GluLysProPro-AsnAspValPhe-----ValThrAlaMetThrGlyAsnPheGlu 486
 Db 2613 CAAAACCAAGGATGATGGGATTTTCTGGATGATAGATACGGCCATGACATTCACAAGTC 2672
 Qy 486 uSerLysGlySerAlaValVal----- 493
 Db 2673 TCCTCAGCGTCTGGAGCAACGACTTTTCTCTTTTAGCGCGCGCCCGCGATTTCTCTA 2732
 Qy 494 -----LeuSerAlaValLeuAspTyArgAsnProAspAsnSerProThrAl 508
 Db 2733 CCGTTTATGCTTCGGTATCAGCTTGAAGTCTGATATCGCAAGGAATCATCAACGAGCAC 2792
 Qy 508 aProLeuTyThrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrAr 528

Db 2793 AACATTGAC-CGGCCGCTTCTTGAGAAAGCTGATG-----GAATTTTGTGATATGACAA 2845
Qy 528 gargpHeG:YProAspArgPhePheGluileLeu----- 539
Db 2846 CAGCTTCGCAAAAGATCGCGCTCGCTCATCTCGAATACAGTCGCGGATGAGTATGCTGG 2905
Qy 539 ----- 539
Db 2906 AAGCGTATCTTCGACCCGATGGAACCTTCAAGGATCATGACAGCTTAGCTATTATTCCTC 2965
Qy 540 ----- 540
Db 2966 AACTAGTTTCATGATACCAACCACTTCGCTGGTTCGTAGGTTCAACCATCACCCCTAC 3025
Qy 545 erThrSerProSerValProProValValSerLysGlnProGlyAlaValGluGluVali 565
Db 3026 TCGTATCTACTTCAGTACCCCTGCG-GTCGAACCCACG----- 3062
Qy 565 leGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheA 585
Db 3063 -----AACCGGCTCATCGCCAGTGGAA-----CATG 3090
Qy 585 lalysAsp-----AlaGlyTyArgLysProLeuArgGluPheGlnLeuArgA 601
Db 3091 CACAAGACTACTTTATCCGCATTCATCAATCACCGATGAAGTTCTTGAAGCAGCATCAAGA 3150
Qy 601 lagLusAspProLysProLlelleLysGluArgValHisPhePheAlaGluThrGlyIleT 621
Db 3151 GCGCGAGAGCGGAACCTCCCTATTCTCGGTGTGTACAGGTAATTGGAGAAAGGGGTG 3210
Qy 621 hrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGluG 641
Db 3211 CCATGGGTCGG-----TGGCATTTGAAGTTCCTTGCCATTTCGCAACTCTCAGATTCGAG 3264
Qy 641 lnArgThrGluPheLysValSerGln-----MetLeuAspT 653
Db 3265 AAGCGGAGCGCTTCATGTTCTGTGAACAGACGAAATTTGACAGCGCATATGATCGAGCCT 3324
Qy 653 rpLeuLeuGlnLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleG 673
Db 3325 GGATGGCGCCGCTTCAGT-----CACATCAAGTTCATTGGCCAAAGTATG 3366
Qy 673 lnLeuGlyLeuSerLysThrTyAlaIleMetThrLeuGluProHisGlnIleArgHisH 693
Db 3367 CGGCTAGACTTGGCGAGTGTTCCTCAACTACCGTTTGTGCTGGCATCCCGGACCGCA 3426
Qy 693 isLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyA 713
Db 3427 GGATGTCACCATCCCTGATGTCGAAAGACGCGATTTTGTCTCACGGATGGAGTAGGCA 3486
Qy 713 rgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProS 733
Db 3487 AGATTTCCCTTTCCTGTCGAAAGATCGTAGCGCATGACTGGAGCATTGACCCGCGCCGT 3546
Qy 733 exAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspT 753
Db 3547 CAGCTTACCAAGTTCGCGATGGGAGTTGCAAGGAGTCTTGGTAACCTGGCCTGATGTA 3606
Qy 753 hrGlyAspGluAspTrpIleGluThrTyProSerGlnArgLysTrpGluCysAspPheV 773
Db 3607 AGGGTATGGAG-----GTTCATATTCCGAAGTCGACGAAAGTTCGTGGCCGAATTC- 3659
Qy 773 alaAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGlnLeuLysSerAlaG 793
Db 3660 -----AATGGCTTGTAGTTGTGTCAGA--TGCCTGCGAGTTCCTCAACCGCCA 3702
Qy 793 lyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaAspLysValLysMetA 813
Db 3703 CTCTCAACCGGCAATCATTGCGCTTCTCTCGTCT----- 3737
Qy 813 rgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysH 833

Db 3738 --CTTGGTGTCCCGATCAAGTCTTTGTAGACATGATGGAGCAG-----CAGCTTT 3786
Qy 833 isAlaLeuAsnArgProValGluPheArgGlnTrpValTyArgLysTrpSerSerArgA 853
Db 3787 CTGACTTCAATGGCGGATGGAGGACAAACA-----AAGS 3822
Qy 853 laThrArgVal-----SerHisGlyArgValPropheLeuAlaGlyL 867
Db 3823 CGACGGCATTCGAAACCGTTTCATTGATGAAATCATATGACACCTATCATTCGCG- 3878
Qy 867 euProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysL 887
Db 3879 -----GAGATCGTTGCTCTAC-----GGCTTATGGGTCTC 3909
Qy 887 ysGlnLysTyLeuGlnAspIle-----AlatTrpAspLeuGlnLysArgL 902
Db 3910 AGAGCCCTTTGTTCGACCCCTACTTCAGCTGTGGGATCATGTCGATCAAA- 3962
Qy 902 ysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyIleTyM 922
Db 3963 -----ACGTTGAAGGAGAAAGCGCACTCAATGTCGAAAGAGTGGCTTCGTGCTCG 4014
Qy 922 etIleAlaAspPheTrpGlyValLeuGluAsnGluValHisValGlyPheSerSerL 942
Db 4015 GCTGTGTTGATGAGACTGGCCTCTGAAAGG-----CACATGAAGTCAATTGAAG 4065
Qy 942 ysPheArgAspGluGlu-----GluSerPheThrLeuLeuSerAspCysA 957
Db 4066 ACTGGAAGACGCTGTCAGAGAGCTCCACAGATCTTCTACAGATTCCTCCGAT--G 4122
Qy 957 spVal-----LeuValAlaArgSerProAlaH 966
Db 4123 ATGTCAATGGCGGTACAGGGTATTACTGGAACCTGCTGCTGGCAGAAACCCGCTC 4182
Qy 966 isPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuL 986
Db 4183 TCCATCTCGTGATATCCGTGTGTGGAGCAGTTGATGCTCTGCTGCTGGCATCAA 4242
Qy 986 ysAspValIlellePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyG 1006
Db 4243 GAGACGTGGTGGTTCCTCCCTGACTGGGACAGACAGCGCCACCATGCTGCTGGTG 4302
Qy 1006 lyAspTyArgGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheV 1026
Db 4303 GAGACCTCGATGTTGATGACTTCTTTGTTATTTGGGATCCTTGTCTATC- 4352
Qy 1026 alaAsnAlaGluMetProLeuGluProAspLeuSerArgTyLeuLysLysAspLysThrT 1046
Db 4352 ----- 4352
Qy 1046 hrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrT 1066
Db 4353 -----CCGAAAGAACGCTCCCATC 4371
Qy 1066 yrAspMetIleGlnLysSerPhe-----HisPheAlaLeuGlnProAsnPheLeuG 1083
Db 4372 CCCCAATGATCAGTAGAGCCCATTCGCGGCAAGGAGTTGGCAACCGAACCCGCTAAACA 4431
Qy 1083 lyMetCysThrAsnTyLysGluArgLeuCysTyTrileAsnAsnSerValSerAsnLysP 1103
Db 4432 ACCGTATTACTCTTTGTTGTTGTTATATGAAGTAC-----AACATCTGC 4476
Qy 1103 roAlaIlelleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleV 1123
Db 4477 CGTTGATCGCTCATGCTCATCTGGCG-----ACACGAGATCGGAGGTGGAAGGATGA 4530
Qy 1123 alPheAsnGluAlaSer-----TrpAlaGlnLeuArgArgGluLeuL 1137
Db 4531 AGACCCGAAATGTAAGTGAAGTATCATGTTTGG-----AAGAAATACCTAT 4581
Qy 1137 euGlyGlyAlaLeuSerLeuProAspProMetTyLysSerAspSerTrpLeuGlyArgG 1157
Db 4582 TG-----ACCATTCACCTGACAGGCTCGAGCTCGCT 4614

| | | | |
|----|------|--|------|
| Db | 2411 | CCAAAGTAGAACTATTTCATTCAACCAACTA---CAAAGATCATGTAGAGTGTCGT | 2467 |
| Qy | 529 | ArgPheGlyProAspArgPhePheGluLeuLeuLeuProSerProThrSerThrSerPro | 548 |
| Db | 2468 | AAATTTGGTTCATCCCAATTCATATGGTTAAATTT---GTAATGAA | 2512 |
| Qy | 549 | SerValProValValSerLysGlnProGlyAlaValGluValLeuGlnTrpLeu | 568 |
| Db | 2513 | TCATTTGGAACCACTT--- | 2527 |
| Qy | 569 | ThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPheAlaLysAspAla | 588 |
| Db | 2528 | -----CARGTGTA | 2536 |
| Qy | 589 | GlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIle | 608 |
| Db | 2537 | AATAATAAACATCCACITTCATGAGATAGAATT-----AAACCAATACATT | 2581 |
| Qy | 609 | LysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAspValPheLys | 628 |
| Db | 2582 | -----GAGGTGGTATATTTGGTGGTGAATCTATTTCG | 2617 |
| Qy | 629 | ThrArgSerValValProAlaGluGluProValGluGlnArgThrGluPheLysValSer | 648 |
| Db | 2618 | TATGCTGGT-----AATAGTAATAGTCAGTTAAGA | 2647 |
| Qy | 649 | GlnMetLeuAspTrpLeuLeu-----GlnLeuAspAsnAsnThr-----TrpGln | 663 |
| Db | 2648 | GAATATTTCATCTTGGTTTCGTTCTTAATCAAAATTTGGTACACATACGGTTAAATTTCCGAGT | 2707 |
| Qy | 664 | -----ProHisLeu-----LysLeuPheSerArgIleGlnLeuGlyLeuSer | 677 |
| Db | 2708 | GGTATAGACATGTTGTATATGTTAGAAAGTTTTCAGATGATTTGGTTAAATGTTTCA | 2767 |
| Qy | 678 | LysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLysThrAspLeu | 697 |
| Db | 2768 | ACAACGATTCCAACGGTTACATTACCACAGAATCGTATC---TATAGAATCCCAAGATATC | 2824 |
| Qy | 698 | LeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSer | 717 |
| Db | 2825 | -----ACCAGAAATACCATGATTTCACTGAGGGATGCTGGTAATTTGGTCCAGAA | 2875 |
| Qy | 718 | ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGly | 737 |
| Db | 2876 | CTTGCTAAACATTTAAATGAATAATTAATTTTCAGACCATCAACATGTGCATATCAAGTT | 2935 |
| Qy | 738 | ArgPheGlySerAlaLysGlyMetTrpValIleAspValAspThrGlyAspGluAsp | 757 |
| Db | 2936 | AGAAATGCTGGAAATAAAGTATGTTGGTGGTAAC---AATCAAGCACCCGATCCATCT | 2992 |
| Qy | 758 | TrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGln | 777 |
| Db | 2993 | GGTATTATATTAGACCACAGTATGGTGAATTCATCAATCGTTTGTGTGATGAACAT | 3052 |
| Qy | 778 | ArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGln | 797 |
| Db | 3053 | AGAAATAGAGATTTTCAGTGGTT---TCAACAACTTCAAGATGTAATTAATAGACAA | 3109 |
| Qy | 798 | LeuLeuProValLeuGluAsp---ArgAlaArgAspLysValLysMetArgGlnAlaIle | 816 |
| Db | 3110 | GTAATAGTTTACTCTCAACATTTGGGTACACAGATATGTT----- | 3151 |
| Qy | 817 | GlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsn | 836 |
| Db | 3152 | -----TTCTTTTCATTAAGATCATCTATCTAAAT | 3181 |
| Qy | 837 | ArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal | 856 |
| Db | 3182 | -----CAAGTGGCTCAATCGTAAATGATACCAATGCTTCCAAACCAAGCAATCGTT | 3232 |
| Qy | 857 | SerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsn | 876 |

| | | | |
|----|------|---|------|
| Db | 3233 | GAATTC-----TTTCCAGATATTACAGAGGGTGACATT--- | 3266 |
| Qy | 877 | PheLeuMetAsnSerGlyPheAspProLysGlnLysIleuGlnAspIleAlaTrp | 896 |
| Db | 3266 | -----TATCAAGATCCTTACATTCGTAGATTTTAAATT | 3298 |
| Qy | 897 | AspIleuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArg | 916 |
| Db | 3299 | AGTCCTTTATAAATTAATAATGGAGAGAAATCAACAAAATGTCATATAGAGATTAAGAT | 3358 |
| Qy | 917 | SerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValIleuGluGlnAsnGluValHis | 936 |
| Db | 3359 | AGTAGAATGTGTTGGGTGTTGTGATCCACAAATTCATTACCACCAATACTGTATTT | 3418 |
| Qy | 937 | ValGlyPheSerSerLysPheArgAspGluGlu-----SerPheThrLeuLeuSer | 954 |
| Db | 3419 | GTTCAATTTGAAGAGAGAAGATGAAGATGATGATGGTCGTAATAATGTAAGAAAGTAAT | 3478 |
| Qy | 955 | AspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVal | 974 |
| Db | 3479 | GARGATTGTAATGTAATTAATAATCCTTGTTACTCATCCTGGTGATGTAAGATATTTA | 3538 |
| Qy | 975 | ArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLys | 994 |
| Db | 3539 | AAGCGTCTGATTAATTAAGATTAAAGACATTTAAGAAATGTTATAGTTTCTCAACAAA | 3598 |
| Qy | 995 | GlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrp | 1014 |
| Db | 3599 | GGTGACGTACCAACTTTAAAGAAATTCAGGTTTCAGATTAGATGGTGATCTATTTC | 3658 |
| Qy | 1015 | ValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluPro | 1034 |
| Db | 3659 | TTTGTGTACGATAAATCATTGAT----- | 3682 |
| Qy | 1035 | AspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHis | 1054 |
| Db | 3682 | ----- | 3682 |
| Qy | 1055 | GlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHis | 1074 |
| Db | 3683 | -----GGTAATCGTTCAAAATCAGAGACAGCCTAT----- | 3712 |
| Qy | 1075 | PheAlaLeuGlnProAsnPheLeuGly-----MetCysThrAsnTyrLysGluArg | 1091 |
| Db | 3713 | -----TTAGGTGACGAGACCGTTTCAACAAATGATATAAAGGCC | 3751 |
| Qy | 1092 | LeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuVal | 1111 |
| Db | 3752 | -----AATGTTTCAATGATCCA---TTTGTTTATCTCTCAATGTAC | 3790 |
| Qy | 1112 | GlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGln | 1131 |
| Db | 3791 | TCATACAAATGTTGAACGTCAAGACCTTGGTAAATGTATAAT-----TCACAT | 3838 |
| Qy | 1132 | LeuArgGArgGluLeuLeuGlyGlyVala-----LeuSerLeuProAspPro | 1146 |
| Db | 3839 | TTGCAATCTCTGATCTTTTCGGTGCAATTCACGAATTTTCAATTCAAATCTCAAGAA | 3898 |
| Qy | 1147 | MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu | 1166 |
| Db | 3899 | TGTTTTAAAGAA-----ATTGATTTATCCA | 3922 |
| Qy | 1167 | LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMet | 1186 |
| Db | 3923 | AAACTCGTATTT---CATGGCACTATTTCAAAGAAGCAAAATGTTTCG----- | 3967 |
| Qy | 1187 | LysAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSer--- | 1205 |
| Db | 3968 | -----TTAAAAACTAAAGGTTTCACCAATATATGCAAAAGAGAGAAATTCACAGCA | 4018 |
| Qy | 1206 | ---TyrTyr----- | 1207 |
| Db | 4019 | GTTTATTATCAATCGAAAAACCATTTATGGGTAAATGATGATCAATTAATGATCGTTGTT | 4078 |

```

QY 1208 -----ThrPhePheLysGluLeuSer---AspLysSer----- 1217
Db 4079 TATATTGGTGAATTCCTTACCAAAATATTAGTCTGTGATAAAGCAATTTGGTTGATGACTAT 4138
QY 1218 -----ArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGlu 1233
Db 4139 GAAATCTANTTAATAGTCTGTAATCTTATATAGTCAGTATAAATCTCAAGTTCATTCA 4198
QY 1234 ValGluLysGluTyrGlyArg----- 1240
Db 4199 TTGTTAAGACATTATTTCAGCGAGTCTGAAGAATCTATAATGATTGGTTTCTTGATCAA 4258
QY 1241 ---LeuValLysAsnLysGluMetArgAsp-----SerLysAspProTyrPro 1255
Db 4259 GGTGTTTATAGTCAAAAGTTTCAAAAGATATCAAAAGGTGAATGAAATGATTATAC--- 4315
QY 1256 ValArgValAsnGlnValTyrGluLysTyrCysAlaIleThrProGluAlaMetAspLys 1275
Db 4316 ATTAAATTCACAAACTTTTGAAGAAC-----GAGTTTTTAAAGGAA 4357
QY 1276 SerGlyAlaAsnTyrAspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAsp 1295
Db 4358 TTGTTGTAACACATAAAGAGAAATGTTTATTGATTTCATAGAGTAAACATT-----GAA 4411
QY 1296 ArgGluMetAsnThrTyr 1301
Db 4412 AAGAAGGTATCAGCATGG 4429

RESULT 9
LOCUS AR145904 3731 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6218142.
ACCESSION AR145904
VERSION AR145904.1 GI:15109093
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3731)
AUTHORS Wasenegger,M., Riedel,L., Schiebel,W. and Sanger,H.L.
TITLE Nucleic acid molecules encoding polypeptides having the enzymatic
activity of an RNA-directed RNA polymerase (RDRP)
JOURNAL
FEATURES
source
location/Qualifiers
1..3731
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 0.000197 Length: 3731
Score: 297.50 Matches: 189
Percent Similarity: 35.78% Conservative: 137
Best Local Similarity: 20.75% Mismatches: 339
Query Match: 4.02% Indels: 247
DB: Gaps: 35

US-09-913-878A-2 (1-1402) x AR145904 (1-3731)
QY 562 GluGluValIleGlnThrPheThrMetGlyGlnHisSerLeuValGlyArgGlnThrArg 581
Db 1349 GAAGACATAGATAACTTCTTCGTGTT-----TCTTTTGTGATGAGGAGTGGAG 1399
QY 582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
Db 1400 AAAGTATTTCTACAGACTTA----- 1420
QY 602 GluAspProLysProIleIleLysGluArgValHisPheAlaGluThrGlyIleThr 621
Db 1421 ----TTACCAAAA-----GCAAGTACTGGAGTGGT 1447
QY 622 PheArgProAspValPheLys-----ThrArgSerValValProAla 635

```

```

Db 1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGAAAGCGTTTGTAAATGGT 1507
QY 636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTyrLeuLeu 655
Db 1508 GATAAAAAATTGAA-----TTTCTTGCAATTTTCATCGAGCCAGTTG----- 1549
QY 656 GlnLeuAspAsnThrTyr-----GlnProHisLeuLysLeu----- 688
Db 1550 ---CGGATAAATTTCAGTGTGGATGTTTGATCAAGACCTGCGCTTACTGCAAAATGATATA 1606
QY 669 -----PheSerArgIle----- 674
Db 1607 AGAGTTGGATGGGTGATTTTTCGAGATCAAGAAATGTCGCAAAATATGTCGCCAGACTT 1666
QY 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLys 694
Db 1667 GGTCAATCTTTGTTGTTCTCTCCAGAGAGACTTTG-----AGTGTCTTTAGCATGAG 1717
QY 695 ThrAspLeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db 1718 ATTGAAGTTATCCCGATGTAAAGTTCATGGAACCAAGTATGCTCTTTCTGATGAAAT 1777
QY 712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
Db 1778 GGTAAAAATATCTGCTGACTTTGCTCATAGAGTTCGCTCAAAATGTCGCCCTTCAATATACC 1837
QY 732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTyrValIleAspValAsp 751
Db 1838 CCATCTGCTTCCAGATTCGTTATGTTGGTATATAAGGTGTTTGGTGGTGTGTATCCGAT 1897
QY 752 AspThrGlyAspGluAspTyrIleGluThrTyrProSerGlnArgLysThrGluCysAsp 771
Db 1898 TCATCAATGAAG-----TTGTCTTTGAGAAGAGCATGTCGAAATATGAATCAGAC 1948
QY 772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db 1949 AACATA-----AAGTTAGATGTCCTTGGAA-----TGGAGCAAAATATCAGCCT 1990
QY 792 AlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db 1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGCT----- 2026
QY 812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db 2027 ---ACACTTCGAGTGAAGATGAAGTTCTCGAAACAGAACAAAGAGCTGTAGATCAG 2083
QY 832 LysHisAlaLeuAsnArgProValGluPheArgGlnThrValTyrGluSerTyrSerSer 851
Db 2084 CTTGATGCTATCTTGATGATTTCTTGAAGGCACAGAGGCTTTGGAATGATGCT--- 2140
QY 852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db 2141 -----CCTGGAGAGAAC 2152
QY 872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db 2153 ACTAATATCTCAAGCAATGCTTAATCTGTGTATAGCTGATCTGAGCCCTTCTT 2212
QY 892 GlnAspIleAlaThrAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db 2213 TCAATGATGTGCAAACTTCGCGCATCCCAAGTTCGATTTGCGGACTAGATCAAGA 2272
QY 912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheThrGlyValLeuGlu 931
Db 2273 ATATTATTCCAAATGGAAGAACATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332
QY 932 GluAsnGluValHisValGlyPheSer-----SerLysPheArgAspGluGlu 947
Db 2333 TATGGTCAGGTGTTGTTTCAGTTTACTGGTGCTGGACATCGACAGATTTCTGACGATTTA 2392
QY 948 GluSerPhe-----ThrLeuLeuSerAspCysAspValLeu 959

```


Db 2393 CATCCATTATAACAGCAGATCCACCAACAGTAATTCATCTTGAAGGAAATGTGGTT 2452
 QY 960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
 Db 2453 GTTGCAGAAAATCCATGCTTGCATCTCGTGATATTCGTTTAAAGGCTGTAATGTT 2512
 QY 980 ProGluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeu 999
 Db 2513 CGAGGCTGCACACATAGTAGTGTGTTGTTATTCCTCAGAAAGGAAAAGACCTCAT 2572
 QY 1000 AlaLysLysLeuSerGlyLysAspTyrAspGlyAspMetAlaIrpValCysIrpAspPro 1019
 Db 2573 CGCAATGATGTTCTGGGAGTGATTTGGATGGGATATCTACTTGTGTTGCTGGATCAA 2632
 QY 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
 Db 2633 GACATGATCCCCCAAGCAAGTCAGCGGATGGAAATATCTCCAGCACCCAGCATACAG 2692
 QY 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
 Db 2693 TTGGACCATGAT----- 2704
 QY 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGluPro 1079
 Db 2705 -----GTCCAAATCAGGAAGTTGAAGAGTACTTCCACCAATATATTGTGAAT 2752
 QY 1080 AsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCysTyrIle 1095
 Db 2753 GACAGTTTGGGAATCATAGCAATGCCATGTCGTATTTCGACAGACA-----GAA 2803
 QY 1096 AsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuVal 1115
 Db 2804 CTGTATATGCCATGATGATCCATGCAAAAACCTTGCTGAGCTTTTTCATATTCAGTG 2863
 QY 1116 AspGlnSerLysGlnGlyValPheAsnGluAlaSerTyrAlaGlnLeuArgArgGlu 1135
 Db 2864 GACTTCCAAAGACTGGTGTCCCTGCAATACCATCT-----CAGTTGGC----- 2911
 QY 1136 LeuLeuGlyGlyAlaLeuSerLeuProMetTyrLysSerAsp-SerTyrLeuGlu 1155
 Db 2912 -----CCTAAAGAAATACCCAGACTTCATGATGAAG 2941
 QY 1155 yArgGlyGluProThrHisIleAspTyrLeuLysPheSerIleAlaArgProAlaIle 1175
 Db 2942 CCGG-----ACAAAGCAGCTAT 2959
 QY 1175 eAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGlu 1193
 Db 2960 ATCTCAAGAAAGAGTTATTGGAAGCTTTC-AGGAAAGTGAAG----- 3001
 QY 1193 uAspGlyAlaHisPheTyrAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIle 1213
 Db 3002 -GACAAAGCA-----CCTCAGGCTAGCTCTATCGGACCTTCACAAGAGATGT 3048
 QY 1213 eSerAspLysSerArgSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGlu 1233
 Db 3049 TGCAGGAGATCATATGATGCTGATATGGAAGTTGATGATTGAAGATTACATTCACGA 3108
 QY 1233 uVal-----GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
 Db 3109 AGCTTTTGACTACAAACTGAATAT-----GACAAAGCTGGGTAAATTT 3153
 QY 1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTyrCysAlaIleThrPr 1270
 Db 3154 AATGGACTACTATGGCATATAAACACAGAG-----GCTGAAATACTTTAG 3195
 QY 1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp-----Se 1282
 Db 3196 TGTGTCATATTGAAGGCAATCAAAACTTTTACCGCGCAAAAGATGCTGAGGCCATTTAG 3255
 QY 1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTyr-- 1301
 Db 3256 TGTGTGTGTGAGGCGCTTGAG-----AAGGAGGCAAGAGCGCTGTT 3297

QY 1302 -----AlaLeuLeuArgAlaSerThrAlaPheLysLeuTyr 1313
 Db 3298 CRAAGGCGTAATGATATGATGACATGTTACCAAGGCTTCGGCTTGGTACACGTTAC 3357
 QY 1313 rTyrHisLysSer----- 3358
 Db 3358 ATATCATCTCAATATTTGGGTTGCTACAACTCAGGGTTGAAAGAGCTCATTTTCATTAG 3417
 QY 1319 spheValTyrGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArg 1339
 Db 3418 CTTTCCTCGTGTGTTTATGACCACTAATCCAGATTAAGAAAGGACAAAGCAGCTGAACAG 3477
 QY 1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
 Db 3478 GCAGAGTCTCACTTCTCAGGCTCACTGAGTCACAGATAGTG----- 3529
 QY 1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
 Db 3530 -----TTGAAATGAGATTCAGTCGAGCGTTAAG 3558
 QY 1379 pProGluValTyrGluValLeuGlyAspAsp 1389
 Db 3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589

RESULT 10

LERDRP LERDRP 3731 bp mRNA linear PLN 18-DEC-1998
 LOCUS L.esculentum mRNA for RNA-directed RNA polymerase.

DEFINITION V10403

VERSION V10403.1 GI:4038591

RDRP gene; RNA-directed RNA polymerase.

KEYWORDS Lycopersicon esculentum (tomato)

SOURCE Lycopersicon esculentum

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

Schiebel, W., Pelissier, T., Riedel, L., Thalmeir, S., Schiebel, R.,

Kempe, D., Lottepeich, F., Sanger, H.B. and Wassenecker, M.

Isolation of an RNA-directed RNA polymerase-specific cDNA clone

from tomato

JOURNAL Plant Cell 10 (12), 2087-2101 (1998)

MEDLINE 99055198

PUBMED 9836747

REFERENCE 2 (bases 1 to 3731)

Wassenecker, M.

Direct Submission

TITLE Submitted (08-JAN-1997) M. Wassenecker, Max-Planck-inst. fuer

Biochemie, Viridiforschung, Am Klopferspitz 18a,

Planegg-Martinsried, 82152, FRG

JOURNAL Location/Qualifiers

FEATURES 1. .3731

source /organism="Lycopersicon esculentum"

/mol_type="mRNA"

/strain="Rutgers"

/db_xref="taxon:4081"

1. .3731

/genes="RDRP"

184..3538

/gene="RDRP"

/EC_number="2.7.7.48"

/codon_start=1

/product="RNA-directed RNA polymerase"

/protein_id="CAA71421.1"

/db_xref="GI:4038592"

/db_xref="GOA:O3Z858"

/db_xref="SPTREMBL:O9Z858"

/translation="MGKTIQVGFPPYLLSAEVVKSFLEKVTGVTCALEVKSKGGS

RAPAKVQADNISADKIIITLANNRLYFGSSYLKAMEMKDIIVQLRAYVDQMGIILAF

GQISDDKFAVLGSEVSGFGLKKFFLSSGSADYKQLSLENIMQVHLRPYIG

QNAQFLLIQLFGAPRIYKLENSCYSFKFTDDQWRTTDFPPSWIGLSLCLQFR

RGVRLNFBESFFHYAERENNTLTQGTFFVFSQKSALEVNVPPEGISIPYKILFKI

SSLVHGGCIPGALNVYFRLVDRRRNVACIEHAELEKLYVKECCVDPVRLTEOYD
 GYLKGRQPKSPSITLDGLVYVRVLVTPCKYFCGPPEVNVNREVRNVEDIDNRL
 RVSPFVDEWEKLTLLPKASTGSGVATNIYERILSTRKGFVIGBKKEFFLAFSP
 QLRDNVMSFASRLGTLANDIRAMWQPSQIKVAKAARLQSGFSRSTLLYRHE
 LEVTPDVHGTSTVYFSGIGIKISGDFAHVASKGLQYTPSAFQIRYGYKGVYVD
 PSSMKLSIRKMSKYESDNILKDLVIGWSKYQPCVLRQLITLLSTLGVKDEVLQEQ
 KEAVDQDLAILHDSLKAQEALELMSPGENTILKAMLNCGYKPDPAEPFLSMLOTPRA
 SKLLDLRTSRIFIPNGTMMCCDDESETLEVGOVFQFTGAGHGEPSDDLHPNNRS
 STNSNFIKGNVVAKNCPGLDIRVLKAVNRALHMDVDCVFPQKGRPHPNES
 GSDLDGDIYFCWDDQMLPPKQVPMETFPAPSPQLOHDDVTIEVEYFTNIYVNDLS
 KILANHVAFADREDPMAMSPCKKLAEFLSIADVDFPKTVPFASIPQLRPKPYDFM
 DKPKDTSISERVISGKLPRKVKDAPQASSIATFRDVARVSADMEVDFGFDIDP
 AFDTYKNDKGLNMDYIGIKETAEILSGGIMKASKTFRKKADEAISVAVRALRKE
 ARANFKRRNDIDMLPKASAWHTYHPTYNGCYNQGLKRAHFISFFWCYVDQILQIK
 KDKARNRPVLNLSRLAQLSHRLVVK"

ORIGIN

Alignment Scores:

Pred. No.: 0.000197 Length: 3731
 Score: 297.50 Matches: 189
 Percent Similarity: 35.78% Conservative: 137
 Best Local Similarity: 20.75% Mismatches: 339
 Query Match: 4.02% Indels: 247
 DB: 8 Gaps: 35

US-09-913-878A-2 (1-1402) x LERDRP (1-3731)

Qy 562 GluGluValIleGlnThrLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
 Db 1349 GAAGACATAGATAACTTCTTCGTCTT-----TCCTTTGTGATGAGGAGTGGAG 1399
 Qy 582 AlaPhePheAlaLysAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
 Db 1400 AAACGTGATTCTACAGACTTA----- 1420
 Qy 602 GluAspProLysProIleLeuLysGluArgValHisPheAlaGluThrGlyLeuThr 621
 Db 1421 ----TTACCAAAA-----GCAGTACTGGAGTGGT 1447
 Qy 622 PheArgProAspValPheLys-----ThrArgSerValValProAla 635
 Db 1448 GTCAAGCAAAACATCTATGAGAGGATCTTATCAACTCTCGCAAGAGGCTTTGTAATTGGT 1507
 Qy 636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
 Db 1508 GATAAAAATTGGAA---TTTCTTCGATTTTCATCGACCCAGTTG----- 1549
 Qy 656 GlnLeuAspAsnThrTrp-----GlnProHisLeuLysLeu----- 668
 Db 1550 ---CGGATAATTACGTGTGGATGTTTGCATCAAGACCTGGCTTACTGCAATGATATA 1606
 Qy 669 -----PheSerArgile-----GlnLeu 674
 Db 1607 AGAGCTTGGATGGGTGATTTTTCGAGATCAAGATGTCGAAAATATGCTTGCAGACTT 1666
 Qy 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLys 694
 Db 1667 GGTCAATCTTTGGTCTCTCCAGAGACTTTG-----ACGTGTTCTTAGCATGAG 1717
 Qy 695 ThrAspLeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
 Db 1718 ATTGAAGTATTCCCGATGTAAGGTTCATGCAACAGCTATGCTTTCTTCATGGAAT 1777
 Qy 712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
 Db 1778 GGTAAAATCTGGTGACTTCTCATAGAGTGGCTCAAAATGTCGCTTCAATATACC 1837
 Qy 732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
 Db 1838 CCATCTGCTTCCAGATTGTTGGTGGATATAAAGGTGTTGGTGGTGGTGGTGGTGGT 1897
 Qy 752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771

1898 TCATCAATGAAG-----TTGCTTTTGAGAAAGCATGTCGAAATATGAATCAGAC 1948
 Qy 772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
 Db 1949 AACATA-----AAGTTAGATGCTCTTGGGA---TCGAGCAAAATATCAGCCT 1990
 Qy 792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
 Db 1991 TGTATCTTANTCGTCACTGATTAGCTCTTGTCT----- 2026
 Qy 812 MetArgGlnAlaIleGlyAspArgLeuLeuAsnAspLeuGlnArgGlnPheSerGluGln 831
 Db 2027 ---ACACTTGGAGTGAAGATGAAGTTCTCGAACAGCAAGCAAGAAAGAACTGTAGATCAG 2083
 Qy 832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
 Db 2084 CTTGATGCTATCTTCGATGATCTTTGAAGGACAGAGGCTTTGGAATGTGCT--- 2140
 Qy 852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
 Db 2141 -----CCTGAGAGAAC 2152
 Qy 872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
 Db 2153 ACTAATATTCTCAAGCAATGCTTAAACTGTGTTTATAAGCCTGATCTGAGCCCTTCTT 2212
 Qy 892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
 Db 2213 TCAATGATGTGCAAAACCTTCGCGCATCCAAAGTTGCTCGATTGGGACTAGATCAAGA 2272
 Qy 912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
 Db 2273 ATATTATTCCAAATCGAAGAACAAATGATGGGATGTTGGATGAATCCAGAACCTTGGAA 2332
 Qy 932 GluAsnGluValHisValGlyPheSer-----SerLysPheArgAspGluGlu 947
 Db 2333 TATGCTCAGGTGTTCTTCAGTTTACTGCTGCTGGCATGGAGAGTTTCTGACGATTA 2392
 Qy 948 GluSerPhe-----ThrLeuLeuSerAspCysAspValLeu 959
 Db 2393 CATCCATTAAATAACAGCAGATCCACCAACAGTAATTCATCTGAGGGAATGGGT 2452
 Qy 960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
 Db 2453 GTTGCACAAAATCCAGCTTGCATCTCGTGATATTTCGTTGTTTAAAGCTGTAAATGTT 2512
 Qy 980 ProGluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeu 999
 Db 2513 CGAGCCTGACCATGCTAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2572
 Qy 1000 AlaLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
 Db 2573 CCGAATGAATGTTCTGGAGTGAATTTGGATGGGATATCTACTTGTCTGGGATCAA 2632
 Qy 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
 Db 2633 GACATGATCCCGCAAGCAAGTCCAGCGATGAATATCTCCACAGCAACCCAGCATACAG 2692
 Qy 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
 Db 2693 TTGACCATGAT----- 2704
 Qy 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
 Db 2705 -----GTCACAAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTAAT 2752
 Qy 1080 AsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCysTyrIle 1095
 Db 2753 GACAGTTTGGGAATCATAGCAAAATGCCATGTCGTTATTTGCACACAGA-----GAA 2803
 Qy 1096 AsnAsnSerValSerAsnLysProAlaIleLeuLeuSerSerLeuValGlyAsnLeuVal 1115
 Db 2804 CCTGATATGCCCATGAGTATCCATCAAAAACCTTGCTGAGCTCTTTTCAATTGACGTG 2863

QY 525 ---ArgLeuThrArgArgPheGlyPro-----AspArgPhe-PheGlu----- 537
Db 3077 AATCGAATCCTCGGAGTACTCTGACCTGAACGACCGGTTCCTTCGAGTCCCAATTCAC 3136
QY 538 -IleLeuIleProSerProThrSerThrSerProSerValProValValSer----- 555
Db 3137 GACGAGTTACCTTTGGCAAAATCTTTTCATCCAGGATTCACAAAGATGACAAACCTG 3196
QY 556 -----LysGlnProGlyAlaValGluGluValIle 565
Db 3197 TACACCCGGGTACACAGGTCATGCAAAACGGTGTTATCATCGGGGACGGTCATTATATAA 3256
QY 566 -GlnTTPLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPheAl 585
Db 3257 TTCTTGGCGTCA--GCAACTCACAGTTTA--GGGAGAACGGC-GCAATCTTCTTTTG 3309
QY 585 alyAsp-----AlaGlyTyArgLysPProLeuArgGluPheGlnLeuAr 600
Db 3310 TGAGACCGATCATGTCACCTGTCCTCCATCCGCAACTGGATGGGGGACTTCCGG----- 3364
QY 600 gAlaGluAspProLysProIleIleLysGluArgValHisPheAlaGluThrGlyI 620
Db 3365 -----CATATCCGCTCCGTCGCAAGTTTGTCTGGAGATGGGTCA 3405
QY 620 eThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValG 640
Db 3406 GTGTTTCACC-----ACAAACGACGAAGTCAATGCATCAGCATACCA----- 3448
QY 640 uGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAsnAs 660
Db 3449 -----AAATTCGCCAGATTGGTGATATACAGCGCGGAGTGGTGACAA 3492
QY 660 nThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTy 680
Db 3493 CATCTGGAAC----- 3502
QY 680 rAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerPr 700
Db 3502 ----- 3502
QY 700 oSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLy 720
Db 3503 -----TTCACTGATGAGTGGGCAAGATATCGTCTCTTTTGAAG 3543
QY 720 sArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGl 740
Db 3544 AATGATTCAAGCGAGCGGTGATTGGCAGAGCGCTTTCATGCTTCAGATGAGATGGG 3603
QY 740 ySerAlaLysGlyMetTrpValIleAspValAspThrGlyAspGluAspTrpIleGl 760
Db 3604 CGGTGTCAGGGCGGTGCTGCTGATGCTGATGCTGATCCAGCGACGAG-----GTCCA 3657
QY 760 uThrTyProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLe 780
Db 3658 TATACGACCTTCTCAGGAGAGTTTAAAGCCGGTGTAC-----AATGGTCT 3702
QY 780 uGluValArgSerValAlaSerGlnLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPr 800
Db 3703 CGAGATCATCAAG--ACCTCTAGTTCGCCCATGCTACTCTCAATTAAGCAGGTTATACC 3759
QY 800 oValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLe 820
Db 3760 TGTCCTGATCGCCCTGGGTGTGAT-----GAGCGCGTTTTCGTGAGAAT 3804
QY 820 uIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValG 840
Db 3805 GCTCGATGATGAGCTCAAGAGTACGACGAA-----GCCTTA----- 3841
QY 840 uPheArgGlnTrpValTyArgLysTySerSerArgAlaThrArgValSerHisGlyAr 860
Db 3842 -----GCTGATTCAATGAAGCGGTGAATCTTCTGGGTCTCA 3879

QY 860 gValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeu---Me 879
Db 3880 AGTC-----GACGAGAACCAACACATGACCATGCGCGAGAT 3918
QY 879 tAsnSerGlyPheAspProLysGlnLysTyLeuGlnAspIleAlaTrpAspLeuGl 899
Db 3919 GTTCGACACGCTTCATGACATCCGAGAGCGGTCTCTC-----TGGACCTTGCT 3966
QY 899 nLysArg---LysCys-----AspThrLeuLysSerLysLeuAsnIleArgValGl 915
Db 3967 TCGTTTGTGGAGTGTGGGTGTTGAAGAGGTTAAAGCAAAAGTTCCCATCAGCGTCAA 4026
QY 915 yArgSerAlaTyIleTyMetIleAlaAspPheTrpGlyValLeuGluGluAsn----- 933
Db 4027 GAACAGCGCTATGATCTTCGGGGTTCGATGAGACCGGTGTCTAGGGGCCCTCACA 4086
QY 934 -----GluVa 935
Db 4087 AGACACCGAAGCAAAAGGTTCAATAACATCGAGTCGCTCCCAACAGATCTTCTCCAAGT 4146
QY 935 lHisVal---GlyPheSerSerLysPheArgAspGluGluSerPheThrLeuLeuSe 954
Db 4147 CCTATTCAGGGGCGATGATGGGAAGAGCACCAAGATTACGAGGTTCATCATCGAATT-- 4204
QY 954 rAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVa 974
Db 4205 -----TGGTAGTAGGCCGTAAACCAGCGCTTCATCTCGGAGATGTGCGGTGT 4254
QY 974 lArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLy 994
Db 4255 TGAGCGAGTTGACGTGCCAGATTACGACATCTCAAGAACGTGCTGCTCTCCAAAGAC 4314
QY 994 sGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyAspGlyAspMetAlatr 1014
Db 4315 TGGAGACGAGGAGTTCGAGCATGTCTCCGTGGTGATGATGACGCGGAGATGACTACTT 4374
QY 1014 pValCysTrpAspProGluIleVal-----AspGlyPh 1025
Db 4375 TGTTTACTGGAGCAGCGACTCATCCAGCGAGTGGGACCCACCTCTGACACACGA 4434
QY 1025 eValAsnAlaGluMetPro-----LeuGluProAspLeuSe 1037
Db 4435 TCGGATTATCGACTCTCGGTGGACAGCCCGCAGATGTGCACAAACGGGACGCTCAC 4494
QY 1037 rArgTyLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGl 1057
Db 4495 TAGGTTCT-----TTGCCAGTATATGAAGAACGATTCCTCTGG 4533
QY 1057 ySerAlaAlaLysGluGlnThrThrTyAspMetIleGlnLysSerPheHisPheAlaLe 1077
Db 4534 CCGTATCGCC-----ACAGCGACCTTTGCTCA 4560
QY 1077 uGlnProAsnPheLeuGly-----MetCysThrAsnTyTrLysGluArgLe 1092
Db 4561 GGCAGATCAGTTGCGAGGGCGGTCAAGCACCCCAAAATGTAAATAATCGCTACGACCTT 4620
QY 1092 uCysTyTrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGl 1112
Db 4621 GTGCTGSC--CGGGGTGAACCTGACCGGTGAAGTAGGCATCGAGCTCGCAGCTCCACTC 4678
QY 1112 yAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLe 1132
Db 4679 AAAAGCTGTCGACTATATCAAGTCCGC-----AAACCCGCGGTCA 4720
QY 1132 uArgArgGluLeuLeuGlyAlaLeuSerLeuProAspProMetTy-----LysSe 1150
Db 4721 GAAGAGACATCTT-----CAGCCTCGAAACTGGCCGCTATGGATGGAGGGGCAAGAA 4774
QY 1150 rAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyTrLeuLysPheSerIl 1170
Db 4775 GTCTCTTACCATCCAGAGCGCGCTGGGTCAAAATTACGACCGGATCAAG----- 4826
QY 1170 eAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLy 1190

| | | | |
|----|------|--|------|
| Qy | 629 | ThrArgSerValValProAlaGluGluProValGluGlnArgThrGluPheLysValSer | 648 |
| Db | 3128 | TATGCTGGT-----AAATAGTAATAGTCAGTTAAGA | 3157 |
| Qy | 649 | GlnMetLeuAspTrpLeu-----GlnLeuAspAsnAsnThr-----TrpGln | 663 |
| Db | 3158 | GAATATTCACTCTGGTTGGTTCTAATCAAAATGGTACACATACGGTTAAAAATTGGAGT | 3217 |
| Qy | 664 | -----ProHisLeu-----LysLeuPheSerArgIleGlnLeuGlyLeuSer | 677 |
| Db | 3218 | GGTATAGAGCATGTTGATAATAGTTAGAAAAGTTTTTCAGATGATTGGTTAATGTTTTTCA | 3277 |
| Qy | 678 | LysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeu | 697 |
| Db | 3278 | ACAAACGATTCCAACAGATTACATACCACAGAAATCGTATC-----TATAGAATCCAAGATATC | 3334 |
| Qy | 698 | LeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSer | 717 |
| Db | 3335 | -----ACCAGAAATACCCATGAA-----TTCACCGAGGGATGGTGAATTTGTCAGAA | 3385 |
| Qy | 718 | ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGly | 737 |
| Db | 3386 | CTTGCTAAACATTATAATGAAATTAATAATTCAGACCACATCAACATGTCATCATCAAGTT | 3445 |
| Qy | 738 | ArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAsp | 757 |
| Db | 3446 | AGAAATAGGTGGAATAAAGTGTGGTGGTGAAC-----AATCAACACCCGATCCATCT | 3502 |
| Qy | 758 | TrpIleGluThrTyrProSerGlnArgLysTrpGlu-----CysAspPheValAspLysHis | 776 |
| Db | 3503 | GGTATTATATAGACCAAGTATGGTGAATTCATCCAAATCGATTGGTGATGAACAT | 3562 |
| Qy | 777 | GlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeu | 796 |
| Db | 3563 | ---AGAACTTAGAGATTTCAGTGTT---TCGACAACTTCAAGATGTAAATTAATAGA | 3616 |
| Qy | 797 | GlnLeuLeuProValLeuGluAsp---ArgAlaArgAspLysValLysMetArgGlnAla | 815 |
| Db | 3617 | CAAGTCATCAGTTTACTCTCAACATTGGGTACACAAGATAATGTT----- | 3661 |
| Qy | 816 | IleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeu | 835 |
| Db | 3662 | -----TTCTTTGCATTACAAGATCATTTATCTA | 3688 |
| Qy | 836 | AsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArg | 855 |
| Db | 3689 | AAT-----CAAGTGGCTCAAAATCGTAATGATACCAATGCTTCCAAACAAGCAATC | 3739 |
| Qy | 856 | ValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeu | 875 |
| Db | 3740 | GTTGAATTC-----TTTCCAGATATTACAGAGGGTGAACTT | 3775 |
| Qy | 876 | AsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAla | 895 |
| Db | 3776 | -----TATCAAGATCCTTACATTCGTAGAAATTTTA | 3805 |
| Qy | 896 | TrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGly | 915 |
| Db | 3806 | ATTAGTCTTTATAAAATTAATAATGGAGAGAAATCAACAAAAATGCTATATAGAGATAAA | 3865 |
| Qy | 916 | ArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGlnAsnGluVal | 935 |
| Db | 3866 | GATAGTAGAATGTTTGGGTGTTTGTGATCCCAAAATTCATTACACCAAAATACTGTA | 3925 |
| Qy | 936 | HisValGlyPheSerSerLysPheArgAspGluGlu-----SerPheThrLeu | 952 |
| Db | 3926 | TTTGTTCAATTGGAAGAGGAAGATGAAGATGATGATGATGATGATGATGATGATGATGAT | 3985 |
| Qy | 953 | LeuSerAspCysAspValLeuAlaAlaArgSerProAlaHisPheProSerAspIleGln | 972 |
| Db | 3986 | GTAATCAAGGATTGGTAATGGTAATTAATAAATCCCTTGTACTCATCTCGTGTGATGAAGA | 4045 |
| Qy | 973 | ArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSer | 992 |

| | | | | |
|------|----------|--------------------------|---|------|
| 4046 | TATCTA | AAAGCGCTGTTGAT | AATAATAAGACATTTAAGAAATGATTTGGTTTCTCTCA | 4105 |
| 993 | ThrLys | GlyAspValPro | LeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMet | 1012 |
| 4106 | ACA | AAAGGTCACGTA | CCAAACTTTAAAGAAATTCAGGTTTCAGATTTAGATGGTGATCGT | 4165 |
| 1013 | AlaTrp | ValCysTrpAspPro | GluIleValAspGlyPheValAsnAlaGluMetProLeu | 1032 |
| 4166 | TATTTCT | TTTGGTTACACATAAT | CATTGATTTGGTGAATCGTTCAACAATCAGAG | 4216 |
| 1033 | GluPro | AspLeuSerArgTyr | Leu | 1044 |
| 4217 | ----- | ACAGCCTATTAGTT | TCGACGCCGTTTCAACAATGATAAAAGGCC | 4264 |
| 1045 | ThrThr | PheLysGln | -----LeuMetAlaSerHisGlyThrGlySerLalaLys | 1061 |
| 4265 | AATGTTT | CAATGATCCATTTGCTTTAT | CCTCAATGCTACTCTACAAATGCTCAAGCTCAA | 4324 |
| 1062 | GluGln | ThrThrAspMetIle | GlnLysSerPheHisPheAlaLeuGlnProAsnPhe | 1081 |
| 4325 | GAA | ----- | CTTGGTAAATTTGATCAT | 4351 |
| 1082 | LeuGly | MetCysThrAsnTyrLys | GluArgLeuCysTyrIleAsnAsnSerValSerAsn | 1101 |
| 4352 | TTGCAAT | CTCTGATCAATAC | ----- | 4384 |
| 1102 | LysPro | AlaIleIleLeuSerSer | LeuValGlyAsnLeuValAspGlnSerLysGlnGly | 1121 |
| 4385 | AAATATT | CAAAATCTCAAAAGAAATGTTT | TAAAGAAATGATTTATCCAAAACTGGT | 4444 |
| 1122 | IleVal | PheAsnGluAlaSerTrp | AlaGlnLeuArgGluLeuLeuGlyAlaLeu | 1141 |
| 4445 | ATTCA | TGGCACTATTCCAAAAGAAAGTAA | CAATTCGGTTCAAACTAGGT | 4495 |
| 1142 | SerLeu | ProAspProMet | -----TyrLysSerAspSer | 1152 |
| 4496 | ---TATCC | ACATTTATATGCAAGAGAAATCA | CCAGAGCTCTATTATCAATCGAAAAAA | 4552 |
| 1153 | TrpLeu | GlyArg | -----GlyGluProThrHisIleIleAspTyrLeu | 1166 |
| 4553 | ATTATGGT | TAAATGATGATCAAAATGATCAGT | TGTTTATATTGGTGATTTCTTTACCA | 4612 |
| 1167 | LysPhe | SerIleAlaArgProAlaIle | AspLysGluLeuGluAlaPheHisAsnAlaMet | 1186 |
| 4613 | AAATATT | AGTCTTGATRAAAGTAATTTGGTT | GTGCTGATGAAATCTATTATAATAGTGCT | 4572 |
| 1187 | LysAla | AlaLysAspThrGlu | AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr | 1206 |
| 4673 | AAAATC | ----- | TTATATAGTCAG | 4690 |
| 1207 | Tyr | ----- | ThrPhePheLysGluIleSerAspLysSerArgSerSerAla | 1221 |
| 4691 | TATAA | CTTCAAGTTTCATTGTTT | AAAGACATTTATTCAGCAGAAATCTGAAGAATCTATA | 4750 |
| 1222 | LeuLeu | PheThrThrLeuLysAsnArgIle | GlyGluValGluLysGluTyrGlyArgLeu | 1241 |
| 4751 | ATG | ----- | ATTGGTTTCTTGATCAAGTTTC | 4777 |
| 1242 | ValLys | AsnLysGluMetArgAsp | -----SerLysAspProTyrProValArg | 1257 |
| 4776 | ATTAGT | CAATAAGTTTCAAAAGATAT | CAAAAGGTGAATGAAAAATCATTTAC | 4834 |
| 1258 | ValAsn | GlnValTyrGluLysTrpCys | AlaIleThrProGluAlaMetAspLysSerGly | 1277 |
| 4835 | ATTCA | CAACAACTTTTGAAAT | -----GAGTTTTAAAGGAATTTGGT | 4876 |
| 1278 | AlaAsn | TyrAspSerLysValIle | ArgLeuLeuSerPheLeuAlaAspArgGlu | 1297 |
| 4877 | GAA | CAACACAAAGAGAAATTC | TTTATTGATTCATAGAGTAAACATT | 4930 |
| 1298 | MetAsn | ThrTrp | 1301 | |

```

Db      4931 GTTTCAGCTGG 4942
RESULT 13
AY148431          3807 bp  mRNA  linear  PLN 12-NOV-2002
LOCUS   Arabidopsis thaliana RNA-dependent RNA polymerase 1 mRNA, complete
DEFINITION
cds
ACCESSION   AY148431
VERSION     AY148431.1  GI:24935274
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 3807)
Yu, D., Fan, B., MacFarlane, S.A. and Chen, Z.
ANALYSIS of the involvement of an inducible Arabidopsis
RNA-dependent RNA polymerase in antiviral defense
Unpublished
JOURNAL
REFERENCE   2 (bases 1 to 3807)
AUTHORS    Yu, D., Fan, B., MacFarlane, S.A. and Chen, Z.
TITLE      Direct Submission
JOURNAL    Submitted (05-SEP-2002) Botany and Plant Pathology, Purdue
University, West Lafayette, IN 47907, USA
FEATURES
source
    /organism="Arabidopsis thaliana"
    /mol_type="mRNA"
    /db_xref="taxon:3702"
    128..3451
    /note="RDRPI"
    /codon_start=1
    /product="RNA-dependent RNA polymerase 1"
    /protein_id="AA064409.1"
    /db_xref="GI:24935275"
    /translation="MGKTIQVFPNGVSAEVEKFLRLTGGSTVVAIKVRQPKGG
    PRVYAIQVTSRHLRLITAAEELYGRSLYKAFVEQDIVPKPRASLHTISGLKM
    FFGQVSTKFTLMSADVCVFGIGMKLHFSFYQKDYRLSELSYENIWQDLHS
    PGRGSAFTVIGAPKIFKEKQDQINLLFGIMDFYSGDGEQWIRTTDTFTSSCHG
    QSTAFCLPVLHVPDRENFANVAEHRASSFLIESGSSYSSNANTLVPVDPDPGG
    SLPPLEFLKNTLVQNACLSGPALDLDFYLLNQKXDRLALDLCLEKFLHGGCCVE
    PAHWLDEYKWSIGKPLFTSISDGLVMTYKRVQVTPARVTFSGDEVNVRNRLR
    HYSKINPLRFVSFVDELEKVRMSDLSPRSQRTKLRYISVDGIVGDKXP
    EFLASSELRENSAMPADIRITAAHRAWMGDFDIRNVAKYAPRLQGFSFSRE
    TLNVRSDLEVIPOVEIISLGRVYVSDIGIKISAEFAKVRKCGLTFFSPSAQIR
    YGKGVAVDPNSKLSLRKMSKFPSENTKLDLAWMSKVQPCYMNRLITLLSTL
    GYTDVFEKKQKQEVDRDLALITPLEAHEALGLMAPCENTNLKALILCGYKDAEP
    FLSMQLNFRASKLLELTKTRIFISGSRSMWGLDETRILEYGVVVQVSDPMRPGR
    RLTITGPVAKNCPCLHSGDVRVLQAVNPALNHWDCVFPQKGLRPHNECSGSL
    DGIIFYVCWDSLPPTSEPMDYTPETQILDHDVTIEEVEYFANYVINDSLGIIA
    NHTAFADKPLKAFSDPCIELAKFSTAVDPFKTGAAVIPQHLVYKEYPDFMEKPD
    KFTYESKNVIGLFEVXERAPPLISIKSFTLDVASKSYDKDMEDVDGFEYVDEAFYQ
    KANYDFKLGNLDYVGIKTEABILSGIMRWKSKFTKRRDAESIGRAVRLRKETLSL
    FNASREENESAKASAWVHTVHSSYWGLYNEGLNRDHLFSFANCYVDKLVRIKTNL
    GRQRQETLERLDHVRFG"
ORIGIN
Alignment Scores:
Pred. No.:      0.00044      Length:      3807
Score:          290.00      Matches:      174
Percent Similarity: 38.02%      Conservative: 129
Best Local Similarity: 21.83%      Mismatches:  102
Query Match:    3.92%      Indels:       195
DB:              8      Gaps:         33

US-09-913-878A-2 (1-1402) x AY148431 (1-3807)
Qy      614 PhePheAlaGluThrGlyIleThrPheArgProAspValPheLysThr-ArgSerValva 633
Db      1396 TTCTACTCAAGAAGAACCACTTATACGACAGGATATATTCTGTTCTCCGATCGGAT 1455
Qy      633 iProLaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAsp-- 652

```

```

Db      1456 TGTTCATAGCGGATAAGAAGTTTGAGTTTCTCGCTTTTCTTCAGCCAGTTGCGGGAGAA 1515
Qy      653 -----TrrLeuLeu-----GlnLeuAspAerAenThr-----TrrG1 663
Db      1516 CTCTGCTTGGATGCTTTCACCAATAGATCGATTACAGCAGCACATATCAGAGCTTGGAT 1575
Qy      663 n-----ProHisLeuLysLeuPheSerArgIleGlnLeuLeuGlyLeuSerLysThrY 680
Db      1576 GGGTGATTGTTTACCACATATAAGAAATGTAGCTAAATATGCTGCAAGGCTTCGTAATCTT 1635
Qy      680 rAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSer-- 699
Db      1636 TAGC---TCYTCAAGAGACACTTAATGTTAGAGTAGTACAGTGAAGTGAATCTCTGA 1692
Qy      700 -----ProSerGlyThrGlyGluValMetAspAspGlyValGlyArgMetSe 715
Db      1693 TGTTCAGATTATATCTTTAGGCACACGCTATGTGTTTCGATCGAATAGAAAGATATC 1752
Qy      715 rArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal---ProSerAl 734
Db      1753 AGCTGAATTGCTAGAAAGGTAGCGAGGAATGTGTTCTTACAGAGTTTCTCCATCTGC 1812
Qy      734 aValGlnGlyArgPheGlySerAlaLysGlyMet-TrrValIleAspValAspAspThrG 754
Db      1813 TTTTCAATCCGTTATGCGCGTTTATAAAGGAGTGGTGGCTGTG----- 1856
Qy      754 lYAspGluAspTrrIleGluThrTyrProSerGlnArgLysTrrGluCysAspPheValA 774
Db      1857 -----ATCC 1860
Qy      774 sPlyHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyL 794
Db      1861 AAATCATCAAGAAA-CTGTCTCTGAGGAAGAGTAGAGCAAAATTCGAATCGAGAACAA 1919
Qy      794 euanLeuGlnLeuLeu-----ProValLeuGluAspArgAlaArgA 808
Db      1920 CCAAGCTTGTATGTTCTGGCGTGGAGCAAGTACCACCTTGTATATGAACAGACAACTGA 1979
Qy      808 sPlyValLysMetArgGlnAlaIleGlyAspArgLeuAenAspLeuGlnArgGlnP 828
Db      1980 TTACACTTTTGTCTACTCTTGGAGTTACGGACAGTGTGTTGAGAGAACAAAGGAAG 2039
Qy      828 heSerGluGlnLysHisAla---LeuAsnArgProValGluPheArgGlnTrrValTyrG 847
Db      2040 TTGTGATCGTCCTCGACGCAATCTTGACTCATCTTTGGAA----- 2080
Qy      847 luserTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyL 867
Db      2081 -----GTCACGAGGCTCTTGGTTTAATGCT----- 2107
Qy      867 euProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysL 887
Db      2108 --CCAGGGGAAAAACAAAATATTCTCAAGGCATTGATCTTGTGTGCTATAAACCTGACG 2165
Qy      887 ysgLnlLysTyrLeuGlnAspIleAlaTrrAspLeuGlnLysArgLysCysAspThrLeu 907
Db      2166 CTGAACCTTTCTCTTCAATGATGCTTCAGAATTTTCAGGCATCTAATGTTGTTGGAACATC 2225
Qy      907 ysSerLysLeuAsnIleArgValGlyArgSerAlaTrrIleTrrMetIleAlaAspPhe 927
Db      2226 GGACAAAAACAAGATTTTCATTTCTGTTGGTGAAGATCTATGATGGATGCTAGACAGA 2285
Qy      927 rPglValLeuGluGlnAsnGluValHisValGlyPheSerSerLysPheArgAspGluG 947
Db      2286 CCAGAAACGCTGGAGTAGTGTGTCAGGTAGTAGTCAGTATTACAGATCCCATGAGCCGGGAA 2345
Qy      947 luGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisP 967
Db      2346 GGCGA-----TTCAATCATCCCGACCTGTGTGTGTTGCCAAAAACCCATGCTGCTGC 2396
Qy      967 heProSerAspIleGlnArgValAlaValPheLysProGluLeuHisSerLeuLysA 987

```


| | | | | | | |
|----|------|--|--------------------|--------------|-----------------------------|------|
| Qy | 843 | GlnTrpVal | TrpGluSer | TrpSerArgAla | ThrArgValSerHisGlyArgValPro | 862 |
| Db | 2107 | GATGTGGCATTTGAAGTTCTTCACGGCATCATGTGCT | ----- | ----- | ----- | 2142 |
| Qy | 863 | PheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsn | PheLeuMetAsnSerGly | 882 | | |
| Db | 2143 | -----GACAGGGGAAACACTCGACCTATCATGCTTAGCGAGT | ----- | ----- | ----- | 2181 |
| Qy | 883 | PheAspProLysGlnLysTrpLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys | 902 | | | |
| Db | 2182 | TTCAAAACCAAAACACGACGGCATCTACGCGGATGTTGTCTTCAGTCAGAAATGCACAA | 2241 | | | |
| Qy | 903 | CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTrpIleLeuMet | 922 | | | |
| Db | 2242 | CTCTGGGGTCTCAGAGAAAATCTCTGTATTTTGTACTTCAGAAAGGTGGCTAATGGGT | 2301 | | | |
| Qy | 923 | IleAlaAspPheTrpGlyValLeuGluGlnGluValHisValGlyPheSer | 940 | | | |
| Db | 2302 | TGCCTAGACGAACGACGGGATCTGAACATGGCCAAATGCTTTATTCAGGTCTCTAAACCG | 2361 | | | |
| Qy | 941 | -----SerLysPheArgAspGluGluGluSerPhe | 950 | | | |
| Db | 2362 | TCTATAGAAAATGTTTCTCCAAACATGGGTCTCGTTTAAAGGAGACAAAGACAGATCTG | 2421 | | | |
| Qy | 951 | ThrLeuLeuSerAspCysAspValLeuAlaAlaArgSerProAlaHisPheProSerAsp | 970 | | | |
| Db | 2422 | GAAGTAGTTAAAGGCTAT--GTAGCCATGTGTAAAGATCCCTTGTCTTACCACGGGGAT | 2478 | | | |
| Qy | 971 | IleLeuArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle | 990 | | | |
| Db | 2479 | GTAAGGATTTAGAACCTGTGTATGTATACCCACCTGCATCATGTATGACTGCCTTATT | 2538 | | | |
| Qy | 991 | PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerClyGlyAspTrpAspGly | 1010 | | | |
| Db | 2539 | TTCCTTCAGAAAGGTGATAGCGCGCATACAACGAAGCTTCTGGCAGTGCACCTGACGGG | 2598 | | | |
| Qy | 1011 | AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet | 1030 | | | |
| Db | 2599 | GACCTGTACTTGTGGCTTGGGATCAGAAACTCATC----- | 2634 | | | |
| Qy | 1031 | ProLeuGluProAspLeuSerArgTyr-----LeuLysLysAspLysThrThrPheLys | 1048 | | | |
| Db | 2635 | -----CCTCCCAACAGGAAAGCTATCGCGCATCATATGATCAGCTGAGAGAGAAG | 2688 | | | |
| Qy | 1049 | GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrTyr----- | 1066 | | | |
| Db | 2689 | AGTTTA-----GGCCGTGCTCTCAACCAACCCAGGTCAA-TATCTACTT | 2729 | | | |
| Qy | 1067 | -----AspMetIleGlnLysSerPhe | 1073 | | | |
| Db | 2730 | CCCCACTTTTCTGTTAGCTGAGATGATGACGCTTAAGATAAAGTACAGAGTGAATC | 2789 | | | |
| Qy | 1074 | HisPhe-----AlaLeuGlnProAsnPheLeuGlyMetCysThr | 1086 | | | |
| Db | 2790 | CAGTATAACTCTCTGATATTTCTGGAAGTGGTGGAAACTTGTGTTGGATGCTTGTAGT | 2849 | | | |
| Qy | 1087 | AsnTyr-----LysGluArgLeuCysTyr | 1094 | | | |
| Db | 2850 | GATTACTTTTGTGCTTCTGAGTTCATAGATTGCTGGCATAGGTTGTTAAACAGCATAC | 2909 | | | |
| Qy | 1095 | IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu | 1114 | | | |
| Db | 2910 | TGTTGTTCACTCTGTAGCAGCTAAG--ATGATGCACATTTAAGAAAACTTTGTTCATATATT | 2966 | | | |
| Qy | 1115 | ValAspGlnSer-----Lys | 1119 | | | |
| Db | 2967 | GTGATCTCGATTTGTATTTTTCGATGCGCTTACTGTGACTGTGCTTGTACTTACTAT | 3026 | | | |
| Qy | 1120 | GlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu----- | 1135 | | | |
| Db | 3027 | CAAGGCAGTATCATGATGTAGTACTTCTGCAATGTAGTTATGAAATCTTTGTATT | 3086 | | | |

QY 1136 -----LeuLeuGlyGly---AlaLeuSerLeuProMet 1147
 Db 3087 TTTTCTGTGCTTGCAGAGTCTGGTAATGATTGCTTTTGTATCAGGAC----- 3140
 QY 1148 TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleLeuAspTyrLeuLys 1167
 Db 3141 -----ATATCGATTTC----- 3152
 QY 1168 PheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 1187
 Db 3153 -----TTTGCAAGAACTTGGCGATGACGATTTGGGCACAAATTGCAAT----- 3197
 QY 1188 AlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyr 1207
 Db 3198 -----GCACACGTCGTTTCATGCTGTGATAGTAGAGTGAGTATGGA 3233
 QY 1208 ThrPhePheLysGluLeuSerAspLysSerArgSerAlaLeuLeuPheThrThrLeu 1227
 Db 3234 GCCATGACGACGAGAA-----TGTTTGTACTGTCGACGAACTA 3269
 QY 1228 LysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal----- 1242
 Db 3270 GCTGCCACTGCAGTGTATCCCAAGACA-----GGGAAATTTGTGTCATGCCCTTCCAC 3326
 QY 1243 ---LysAsnLysGluMetArgAsp-----SerLysAspProTyrProVal----- 1256
 Db 3327 CTAAACCAAACTCTACCCAGATTTTCATGGGAAAGAGACTACCAAACTTACAACTCG 3386
 QY 1257 -----ArgValAsnGluValTyrGluLysTrpCys--- 1266
 Db 3387 AACAAATCTTGGTGGCTTTACAGCGGTTAAAGAGTTTATGATGAAGATCGAGAA 3446
 QY 1267 AlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArg 1286
 Db 3447 GCTTCTCAGAGAAAGACAGACAGACCAAGTGCATCCCTTATGACGCT----- 3494
 QY 1287 LeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp 1301
 Db 3495 GTTCTTGAATACCGGATTGAAGATTGATCCCTGAGGCGATGG 3539

RESULT 15
 AX204839
 LOCUS AX204839 6863 bp DNA linear PAT 30-AUG-2001
 DEFINITION Sequence 1 from Patent WO0155407.
 ACCESSION AX204839
 VERSION AX204839.1 GI:15394183

KEYWORDS Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1
 AUTHORS Beclin,C., Elmayan,T., Mourrain,P. and Vaucheret,H.
 TITLE Novel sg2 plant gene and use thereof
 JOURNAL Patent: WO 0155407-A 1 02-AUG-2001;
 Rhobio (FR) : INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)

FEATURES
 source
 1..6863
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.00098 Length: 6863
 Score: 289.50 Matches: 190
 Percent Similarity: 35.43% Conservative: 120
 Best Local Similarity: 21.71% Mismatches: 297
 Query Match: 3.91% Indels: 269
 DB: 6 Gaps: 41

US-09-913-878A-2 (1-1402) x AX204839 (1-6863)

QY 535 PhePheGluIleLeuLeuProSerProThrSerThrSerProSerValProProValVal 554
 Db 2245 TTTTGTCTCTATTTGAGAGGATTGCAAGTTCTTTCTCCAGAGACCTACGTT 2304
 QY 555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
 Db 2305 TTTAAAGAGTGAAGACATA-----TTAACCGATGGG-----TTTAAA 2343
 QY 575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
 Db 2344 CTATGTGGTGAAGAAATAC---AGTTTCTAGCAATCTTCAGCCAAAT----- 2385
 QY 595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleLeuLysGluArgValHisPhe 614
 Db 2386 -----CAACTGAGA-----GACCGCTCGCATGCTTC 2412
 QY 615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
 Db 2413 TTTGTCTGAAGACGGG-----AAAAACAGCT----- 2436
 QY 635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
 Db 2437 -----GTGTCAGATATAAAACATGGATG 2460
 QY 655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
 Db 2461 GGGAGAGTCAAGACAAAGATGTG---GCAAAATGTGCTGTAGGATGGGC 2508
 QY 674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
 Db 2509 CTGTGCTTCTCCCACTATGCACTGTAGATGCTATGCTCAGGAGTT----- 2559
 QY 694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValArg 713
 Db 2560 GACACTGAGGTTCAGATATTGAGAGAAATGGGTATGTTTCTCTCAGCGAATTGTGACA 2619
 QY 714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal----- 731
 Db 2620 ATCACCTGACCTCGCTGACGAGTAATGAGAACTTAAGTTG---GATGTGCACTAC 2676
 QY 732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
 Db 2677 AGCCCTTGTGCTTATCAGATACGTTACGAGGTTTCAAGGGGTT----- 2721
 QY 751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln----- 765
 Db 2722 -----GTTGCTCGTGGCCATCAAAAGTGTGAATCAGG 2757
 QY 766 ---ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
 Db 2758 CTAGCCCTTCGAGACACTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 2814
 QY 783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeu 802
 Db 2815 TGTTCCT---TGGACGAGGTTTCAACCTGGGTTCTTAATCGGAGATAATACCCCTCTA 2871
 QY 803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
 Db 2872 TCC-----GTACTAGTCTTCGGATGAATATCTGG 2904
 QY 823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
 Db 2905 GATATGCAG-----GAATCCATGCTCTATAAACTGAACCCGATCCTCTGTATGATACA 2955
 QY 843 GlnTrpValTyrGluSerTyrSerArgAlaThrArgValSerHisGlyArgValPro 862
 Db 2956 GATGTGGCATTTGAAGCTTCTCAGCGCATCATGTGCT----- 2991
 QY 863 PheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGly 882
 Db 2992 -----GAACAGGGAACACTGCAGCTATCATGCTTAGTCAGGT 3030

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 02:00:42 ; Search time 833.391 Seconds
(without alignments)
7446.671 Million cell updates/sec

Title: US-09-913-878A-2

Perfect score: 7397
Sequence: 1 MNPTPRKNSFVEIINRL.....YEVLGDDDFDGIPTGNGDY 1402

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-Q/cn2_1/USPTO spool/US09913878/runat 29032004 082007 8357/app query.fasta_1.2254
-DB=N Geneseq 29Jan04 -QFMT=fas+ap -SUFFIX=ring -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878 @CN 1 1 797 @runat 29032004 082007 8357 -NCPU=6 -ICPU=3
-NO WAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980a:.*
2: Geneseqn1990a:.*
3: Geneseqn2000a:.*
4: Geneseqn2001a:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002a:.*
7: Geneseqn2003a:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004a:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 7397 | 100.0 | 8045 | 3 AAA65171 | AAA65171 Neurospor |
| 2 | 341 | 4.6 | 615 | 7 AB254843 | AB254843 Aspergill |
| 3 | 328 | 4.4 | 3591 | 4 AAH77704 | AAH77704 Nucleotid |
| 4 | 319.5 | 4.3 | 3901 | 3 AAC63740 | AAC63740 Maize RNA |
| 5 | 317 | 4.3 | 2816 | 3 AAC63741 | AAC63741 Rice RNA |
| 6 | 297.5 | 4.0 | 3731 | 4 AAS17845 | AAS17845 Tomato RN |
| 7 | 297.5 | 4.0 | 3731 | 4 AAS17837 | AAS17837 Tomato RN |
| 8 | 297.5 | 4.0 | 3731 | 4 AAD04370 | AAD04370 Tomato RN |

| | | | | | |
|----|-------|-----|--------|------------|--------------------|
| 9 | 289.5 | 3.9 | 6863 | 4 AAH77703 | AAH77703 Nucleotid |
| 10 | 283 | 3.8 | 3807 | 3 AAC63742 | AAC63742 Soybean R |
| 11 | 267.5 | 3.6 | 3737 | 3 AAC63739 | AAC63739 Maize RNA |
| 12 | 236 | 3.2 | 1281 | 3 AAC63738 | AAC63738 Maize RNA |
| 13 | 195.5 | 2.6 | 4551 | 7 ACA42620 | ACA42620 Prokaryot |
| 14 | 192 | 2.6 | 49377 | 2 AAU05287 | AAU05287 The sorap |
| 15 | 189.5 | 2.5 | 4148 | 3 AA290112 | AA290112 Human ant |
| 16 | 189 | 2.5 | 4010 | 6 ABI99508 | ABI99508 Mouse isc |
| 17 | 185.5 | 2.5 | 5922 | 4 AB118812 | AB118812 Drosophil |
| 18 | 185 | 2.5 | 5710 | 4 AB118813 | AB118813 Drosophil |
| 19 | 184 | 2.5 | 3824 | 4 AA157835 | AA157835 Human pol |
| 20 | 183 | 2.5 | 3813 | 3 AA290111 | AA290111 Human ant |
| 21 | 178.5 | 2.4 | 28958 | 2 AAT89956 | AAT89956 Sorangium |
| 22 | 178.5 | 2.4 | 28958 | 3 AA752399 | AA752399 DNA seque |
| 23 | 176 | 2.4 | 4868 | 7 ABT19128 | ABT19128 Aspergill |
| 24 | 176 | 2.4 | 4733 | 7 ABT18534 | ABT18534 Aspergill |
| 25 | 176 | 2.4 | 5439 | 7 ABT20948 | ABT20948 Aspergill |
| 26 | 176 | 2.4 | 5550 | 7 ABT20350 | ABT20350 Aspergill |
| 27 | 176 | 2.4 | 6733 | 7 ABT17940 | ABT17940 Aspergill |
| 28 | 176 | 2.4 | 7550 | 7 ABT19754 | ABT19754 Aspergill |
| 29 | 175 | 2.4 | 13329 | 5 AA871562 | AA871562 DNA encod |
| 30 | 174 | 2.4 | 2469 | 7 ACA38023 | ACA38023 Prokaryot |
| 31 | 174 | 2.4 | 14770 | 6 AB161893 | AB161893 Colon ade |
| 32 | 174 | 2.4 | 14770 | 6 AB168257 | AB168257 Kidney ca |
| 33 | 174 | 2.4 | 14770 | 6 AB168538 | AB168538 Kidney ca |
| 34 | 174 | 2.4 | 14770 | 9 ADB31314 | ADB31314 Testoster |
| 35 | 174 | 2.4 | 14797 | 5 AA865159 | AA865159 DNA encod |
| 36 | 174 | 2.4 | 15187 | 5 AA861311 | AA861311 DNA encod |
| 37 | 174 | 2.4 | 36535 | 7 AA156705 | AA156705 Chimpanze |
| 38 | 173.5 | 2.3 | 6814 | 4 AB117599 | AB117599 Drosophil |
| 39 | 173.5 | 2.3 | 9059 | 4 AB117598 | AB117598 Drosophil |
| 40 | 173.5 | 2.3 | 10074 | 6 ABK13416 | ABK13416 Drosophil |
| 41 | 173.5 | 2.3 | 28598 | 2 AAT06769 | AAT06769 Sorangium |
| 42 | 172 | 2.3 | 1353 | 8 ACF06132 | ACF06132 Bacterial |
| 43 | 171 | 2.3 | 52397 | 2 AAT51411 | AAT51411 Mycobacte |
| 44 | 171 | 2.3 | 52398 | 2 AAQ47357 | AAQ47357 LS mycoba |
| 45 | 171 | 2.3 | 113193 | 7 AAD54645 | AAD54645 Streptomy |

ALIGNMENTS

RESULT 1

AAA65171
ID AAA65171 standard; DNA; 8045 BP.

XX
AC AAA65171;

XX
DT 16-NOV-2000 (first entry)

XX
DE Neurospora crassa qde-1 gene.

XX
KW Gene silencing; quelling deficient; qde-1; al-1; ds.

XX
OS Neurospora crassa.

XX
PH Key Location/Qualifiers

FT CDS 2447..6655

FT /*tag= a

FT /product= "QDE-1"

XX
PN WC2000050581-A2.

XX
PD 31-AUG-2000.

XX
PF 16-FEB-2000; 2000NO-IT0000048.

XX
PR 22-FEB-1999; 99IT-RM000117.

XX
PA (UJVO-) UNIV ROMA LA SAPIENZA.

XX
PI Macino G, Cogoni C;

XX
DR WPI; 2000-579171/54.

DR P-PSDB; AAB13956.
XX Novel polynucleotide encoding a polypeptide which has a silencing
PT activity and comprising a RNA-dependent RNA polymerase domain.
XX
XX Claim 1; Page 31-43; 48pp; English.
XX
CC The present sequence is the Neurospora crassa qde-1 gene. This gene has
CC silencing activity. The qde-1 gene was isolated by mutational analysis of
CC an al-1 transgenic strain. This strain had an albino phenotype resulting
CC from post-transcriptional silencing of the endogenous al-1 gene.
CC Reversion of this phenotype indicated a mutation in a silencing gene. The
CC silencing gene, qde-1, could then be isolated. Modulation of qde-1
CC expression may be used to inactivate genes and to silence suppression of
CC genes
XX
SQ Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 8045
Score: 7397.00 Matches: 1402
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-913-878A-2 (1-1402) x AAA65171 (1-8045)
QY 1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleAAsnArgLeu 20
Db 2447 ATGACCCCTATTACTCTAGGAGAGGAATAGCCCGTCGAGGAATCATTAACCGGCTC 2506
QY 21 AsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThrLeuThrProHisArg 40
Db 2507 AATACGACTACAACTGGGCTCCAGTGTGTCGAGACACAACTCTCACCCGCCACCGC 2566
QY 41 ArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAla 60
Db 2567 CGGAGGAGCTGGCGAGAGTACGAGGATTCGGTGCCTATGACAAAGATCTACAGAGCC 2626
QY 61 LeuAsnPheLeuTyrTrpArgLysAspSerLeuAsnGlnAlaGluAlaAsnPhePhe 80
Db 2627 CTGAACCTTTCTACTGGCGGAGGATGACTCCCTGGAACCGGAGGAGCAACTCTCTC 2686
QY 81 IleGluAlaLysAlaAsnSerAsnTrpValProLysAlaHisAlaAspProAspThr 100
Db 2687 ATCGAGGCCAAAGCTCGAGCTCGAATCGGTGCCCAAGCCACCGCCCTGACACG 2746
QY 101 LeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGln 120
Db 2747 CTTCCGTGTGTCAGAGAACTCCCGCGCCGCTACTGCCGCCCAACAATGGGCAATTGCG 2806
QY 121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
Db 2807 ACTGTGTGCTCGAGGTGCTTAATAGTATTATGCCACCTCCCAATAACACACAGGTCTGA 2866
QY 141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
Db 2867 ACGTTTGGCAGAACTCTAAGCGGCCCAAGTGGCCCTGAGCCGCCCAACCTCTACCAACACC 2926
QY 161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
Db 2927 AAACGCAAGGATGAGCCGCCAATGTCACTTCGTGTATCGGCCCAACAGCTCTGTGACT 2986
QY 181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
Db 2987 CGCTCTGCCACAGGTCCTCTATTACGCGCGCGGATACCCCTAAAGTTCGCCGATCCA 3046
QY 201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
Db 3047 GTGAATACCGGTTCCAAACGACCACTCTCTCGAGAGTGAGAACTCTCAATCAGTGACCAAG 3106
QY 221 ArgAlaLysGlyLysSerAspAsnValAlaAlaAlaAlaProProValProIle 240

Db 3107 CGGGCCAAAGGCAAGCTGTCTGATAATGTTCGCGTTCGCGCGCCCGCGCTGCTATT 3166
QY 241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db 3167 GCAGGCGCTTTGGACAAAGGTACCGACTCGAAGGATGCCAATACGAGAGATCCACGGCG 3226
QY 261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db 3227 ACAGGTCATAGACGAGCGGACCCAGGTGGATTCTCTTTGATACATCTCAAGGCACTTCTAT 3286
QY 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrClnSerSerPheGlu 300
Db 3287 GGTTCAGGTGTCTTCAGCGCTTCGCGTCAATCAGAGCACCACCCAGAGTAGTTTIGAG 3346
QY 301 AlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGly 320
Db 3347 GCTCTCTTTCACAGCCAGAGAGAGCGCTGTGGATGCCACGCTCTTTGAGGCTGGA 3406
QY 321 HisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsnGlnPro 340
Db 3407 CACTTGATTGAGTCTCTTAGCAAGAGAGAAACAACAGTCCACATATACCATACCGGCC 3466
QY 341 LeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
Db 3467 CTTTCATCGTCTCCAGGGTGAACTTCGTTCCAGCACCTTACTATGAGTCGTTTCCAGT 3526
QY 361 SerGlyGlyGluClyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
Db 3527 TCCGGCGGCGAGCGGCCAATTCGAGCGAGTTCGCTCAATAGGAGTGGCTCGGACCGAA 3586
QY 381 GluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsn 400
Db 3587 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCGGTGTTGCAGTCCGCTGAGAAAT 3646
QY 401 IleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluVal 420
Db 3647 ATTGCGCGGAAATTTCCAAATGGGTACACGAGGCTCTCTCGCTGTTTCATCGGGAAGT 3706
QY 421 ThrArgLeuPhePheHisCysLysValAspLeuGluAspGluSerLeuLysTyr 440
Db 3707 ACCAGACTCTTTATGACTGCAAGTAGACTTGGAGAGAGAGAGGCTGGGCTTAAGTAC 3766
QY 441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeu 460
Db 3767 GACCCCTCTCTGTGTACCGCGCGGATGTACAGATATCTGGAAGACTCTCTACCGGCTT 3826
QY 461 AspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThrAla 480
Db 3827 GATGCTTTCCGTGGTAAACCTTTCCAGAAAAGCCGCCCAACAGCTGTGTGTGACCGCA 3886
QY 481 MetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyr 500
Db 3887 ATGACGGGCAACTTTCAGAGCAAAAGGTAGTGGCTGTGTTCTCTCTGCTGTTCTAGACTAC 3946
QY 501 AsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPhe 520
Db 3947 AATCCCGCAACCTCGGCTACTCGCCCTTACCTTTGGAAGCTGAAGCCGCTCACTGTTTC 4006
QY 521 GluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluLeuLeuIle 540
Db 4007 GAGCAGGCTGTCTGCTACCCGCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATA 4066
QY 541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560
Db 4067 CCGTCCCTTACGAGCACCAGCCCAAGTGTACCGCGGTGGTTCAGCAAAACACACGAGTGGC 4126
QY 561 ValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrp 580
Db 4127 GTCGAGAGAGTCAATCAGTGGCTCAGATGGGCAACATCTCTGTTAGCGCCCAATGG 4186
QY 581 ArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArg 600

Db 4187 CCGCGCTTTCTTCGCGCAAGATCGCGGATACAGAAACCTCTCAGGGAGTTCCAGCTCCGC 4246
Qy 601 AlaGluAspProLysProIleLeuLysGluArgValHisPheAlaGluThrGlyLe 620
Db 4247 GCCGAGGACCCAAACCACATCATCAGGAGAGAGTCCACTCTTTTGCAGACCGGCATT 4306
Qy 621 ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
Db 4307 ACGTTCCGACCTGATGTGTTCAGACGAGATCTGTCTCGTCCGGCAGAGAACTGTAGAG 4366
Qy 641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuLysAsnAsn 660
Db 4367 CAACGACCGAGTTCAAAGTTAGTCAAATGTGGACTGGCTCGCTGCAACTCGACAAAC 4426
Qy 661 ThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyr 680
Db 4427 ACTTGGCAGCGCACCTCAAGTTGTCTCCCGTATCCAGCTCGGTCTCAGTAAGACATAT 4486
Qy 681 AlaIleMetThrLeuGluProHisGlnIleArgHisLysThrAspLeuLeuSerPro 700
Db 4487 GCCATTATGACATTGGAGCCCTCACAGATCAGACACCAAGACCGCATCTCTTTTCACCT 4546
Qy 701 SerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLys 720
Db 4547 TCAGGCACTGGCGAAGTGTATGATGACGGTGTAGCGCGCATGTCCGAGAGCGTGGCCAG 4606
Qy 721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
Db 4607 AGGATACGCGATGTTCTCGGTTTGGGTGATGTGCCCTCTGTCTGCAAGGCGGTTTGGT 4666
Qy 741 SerAlaLysGlyMetTrpValIleAspValAspThrGlyAspGluAspTrpIleGlu 760
Db 4667 TCGGCAAGGAGATGTGGTTATCGAGTTGACGACACAGCGCATGAGATTGGATCGAG 4726
Qy 761 ThrTrpProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
Db 4727 ACATACCGCTCCAGCGCAAGTGGATGCGACTTGGTTGATAAATCAACATCAACGTACCTC 4786
Qy 781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 800
Db 4787 GAAGTCCGAGCGTGGCTCTGAACTGAAGTCAGCTGTCTCAACCTACAGCTGTATACCT 4846
Qy 801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
Db 4847 GTCCCTGGAGATAGAGCCAGGACCAAGGTGAAGATGGCCAGGCAATCGGTGACCGTCTT 4906
Qy 821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
Db 4907 ATCAACGATTTGCAACGACAGTTTCAGCGAGCAAAAGCATGCTTTGAAATCGCCAGTGGAA 4966
Qy 841 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 860
Db 4967 TTTCCGCCAATGGGTTTACGAGATTTATCCAGTCCGCGCACTCGAGTCAGCCAGCGCGT 5026
Qy 861 ValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsn 880
Db 5027 GTGCCCTTTCTTGTGGCTACCTGACAGTCAAGAGGAGACACTGAACTTTCTTGATGAAC 5086
Qy 881 SerGlyPheAspProLysGlnLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
Db 5087 AGTGGTTTCGATCCCAAGAACAAAGTACTTGCACAGATCGCTCGGATCTTCAAAG 5146
Qy 901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
Db 5147 CGGAAATGTGACAGTTGAAGTCCAAAGCTGAACATCCGTGTCGTCGATCAGCATACATT 5206
Qy 921 TyrMetIleAlaAspPheTrpGlyValLeuGluAsnGluValHisValGlyPheSer 940
Db 5207 TACATGATTCCGCAATTTCTGGGGTGTCTTGAGGAAATAGAGTTTCATGTCCGATTTCTCC 5266
Qy 941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
Db 5267 TCAAAGTTCAGACGAGGAGGAGTCTTTTACACTCTATCCGACTGTGATGTCTTCGTCG 5326

Qy 961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
Db 5327 GCCTGATCCCGACCCCATTTCCCTAGTGATATCCAAACGGGTTCGAGCAGTCTTCAAGCCA 5386
Qy 981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
Db 5387 GASCTCCAGTCTCAGGATGTATCATCTCTCTACTANAGGAGATGTACCGCTTGCT 5446
Qy 1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
Db 5447 AAGAAGCTATCTGTGTGGAGACTACGACGGCGATATGGCCTGGGTCTCTGGATCCGGAG 5506
Qy 1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
Db 5507 ATCGTTCGATGGTTCGTCATATCGGGAATGCTCTCTGAGACCCGACCTGTCTAGTACCTA 5566
Qy 1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
Db 5567 AAGAAGCAAAACGACTTTCAAACAACTTATGGCCTCACAGGCAACGGCTCAGCGGCC 5626
Qy 1061 LysGluGluThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
Db 5627 AAAGACAGACTACATACGATATGATCCAGAAAGAGCTTCCATTTCCCTCTGAGCCCAAC 5686
Qy 1081 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 1100
Db 5687 TTCTTGGCATGTGCACTAACTACAAAGAAAGCTCTGTGTACATCAACAATAGTGTGTCT 5746
Qy 1101 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 1120
Db 5747 ACAAGCCGGCCATCAATCTTAGTTACCTGGTGGGAACCTCGTCGATCAGAGCAAGCAA 5806
Qy 1121 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyAla 1140
Db 5807 GGTATTGTCTTTAAACGAAGCAAGCTGGGCTCAATTCGTAGGGAACCTGTTGGCGTGCA 5866
Qy 1141 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyClnProThr 1160
Db 5867 TTGTCCCTTCCTGACCCCAATGTACAAGACGACAGTTGGCTCGCGCGCGAGCGCTTACC 5926
Qy 1161 HisIleLeuAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 1180
Db 5927 CACATTTATGACTACTGAAATCTTCATCGCCAGCGCTCGATTTGACAGGAATGGNA 5986
Qy 1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
Db 5987 GCCTTCCACAATGCCATGAAGCGCCCAAGGATACAGAAGACGGCGCTCACCTTTTGGGAT 6046
Qy 1201 ProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSer 1220
Db 6047 CCGGATCTCGTTCCTTCTACACAGTTCTTCAAGGAGATTAGCGCAAGTCGGATCGTCC 6106
Qy 1221 AlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg 1240
Db 6107 GCACCTGCTATTCCAGCTCTGAAGAACCGTATCGCGGAAGTCGAGAAAGAAATATGGCAGG 6166
Qy 1241 LeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGln 1260
Db 6167 TTGGTCAAAACAAAGAGATGAGACAGCAGGACCCCTTACCTCTCCGCGTCAACGAG 6226
Qy 1261 ValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyr 1280
Db 6227 GTTTATGAAAATGTGTGCCCATCACGCTGAGGCGATGGACAATCCGGAGCAAAATAT 6286
Qy 1281 AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
Db 6287 GATTCTAAGGATCAGGTTGTGGAGTGTCTTCTCCGCGGACCGGTGAGATGAATACA 6346
Qy 1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
Db 6347 TGGGCTATTGCTCAGGGCTAGCACGCGCTTTCAAGCTGTACTACCAACAAGAGCCCCAAGTTC 6406

QY 1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrlleLysAlaGlnMetThrSerArgPro 1340
 Db 6407 GTGTGGCAGATGGCGGCAGACAGCTCGGTACATTAAAGGCGCAGATACAGAGACCC 6466
 QY 1341 GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
 Db 6467 GGTGAAGCGCCCGCGGTGTGATACCGGTTCATGTATGCGGGCTTCATGCCGATAAG 6526
 QY 1361 LysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspPro 1380
 Db 6527 AAGTTTACGAAGCAGTATGTGGCCAGGCTGGAGGCGCATGGATCGGATACCTCGATCCG 6586
 QY 1381 GluValTyrGluValLeuGlyAspAspPheAspGlyIleGlyPheThrGlyAsnGly 1400
 Db 6587 GAGGTCTATGAGTGTGGCGCATGTATTTTGATGGAAATGGTTTCACAGGGATGGC 6646
 QY 1401 AspTyr 1402
 Db 6647 GATTAT 6652

RESULT 2
 ABZ54843
 ID ABZ54843 standard; cDNA; 615 BP.
 XX AC ABZ54843;
 XX DT 28-MAR-2003 (first entry)
 XX DE Aspergillus oryzae polynucleotide SEQ ID NO 3956.
 XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
 XX KW expressed sequence tag; gene; ss.
 XX OS Aspergillus oryzae.
 XX PN WO200279476-A1.
 XX PD 10-OCT-2002.
 XX PF 22-MAR-2002; 2002WO-IB000890.
 XX PR 30-MAR-2001; 2001JP-00098371.
 XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX PA (NARE-) NAT RES INST BREWING.
 XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.
 XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
 XX WPI; 2003-046817/04.
 XX FT Detection of expression of specific *Aspergillus* genes for monitoring the
 FT fermentation and growth conditions of the fungus, using DNA probes.
 XX PS Claim 1; SEQ ID NO 3956; 48pp + Sequence Listing; Japanese.
 XX CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridizing
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of *Aspergillus oryzae* which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 615 BP; 151 A; 148 C; 162 G; 154 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,41e-12 Length: 615
 Score: 341.00 Matches: 89
 Percent Similarity: 54.75% Conservative: 32
 Best Local Similarity: 40.27% Mismatches: 80
 Query Match: 4.61% Indels: 21
 DB: 7 Gaps: 6

US-09-913-878A-2 (1-1402) x ABZ54843 (1-615)

QY 935 ValHisValGlyPheSerSerLysPheArg-AspGluGluGluSerPhe-----ThrLe 952
 Db 1 GTTCACCTTGGCTTCTCCAGTAACCTGCTGTATCTCTGAAGGCCAATTCGAAGACAATCT 60
 QY 952 uLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleG 972
 Db 61 ACTCGACGGTATGAGACGTTCTTGTAGCTCGGCTCCCGCACATGTCCGCTCTGATATCCA 120
 QY 972 nArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleLePheSe 992
 Db 121 CGACGGAAGAGCTGTGTGGAAGCCAGAGCTCGGCATTTCAAAGATGTTATCGTTTCCC 180
 QY 992 rThrLysGlyAspValProLeuAlaLysLysSerGlyLysAspTyrAspLysAspMe 1012
 Db 181 TACTGTGGACAAACCCACTGGCGCACATGTTATCCGTGGTGATACGCGGTATAC 240
 QY 1012 tAlaTrpValCysTrpAspProGluLeuValAspGlyPheValAsnAlaGluMet--Pr 1031
 Db 241 AAGCTGGGTTTGTGGGACCAAAATATTGTCCAGAAATTTCCGTAATTCGACCTATCCAC 300
 QY 1031 oLeuGluProAspLeuSerArgTyr--LeuLysLysAspLysThrThrPheLysGlnLe 1050
 Db 301 CATGGAAATACCTCGGAGCACCTTGTGTTTGGAGAACATAACGTGCCAATGAAAGATAT 360
 QY 1050 uMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrTyrAspMetIleG 1070
 Db 361 T-----GACTCATGGGACGAGTTTCTTCA 384
 QY 1070 nLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysG 1090
 Db 385 AAGCAGCTTTTACGTTTCAATTTAACCATGTTCTAATCTGGGAGGTGACTGTGCGAGCAGA 444
 QY 1090 uArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLe 1110
 Db 445 GAAGATATCCTAC---GACGAATCGATCGACTCGCGCAATGCCAAAGAGCTCGCTGTCT 501
 QY 1110 uValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAl 1130
 Db 502 CCTCAGTCATTTGGTGACGCTCGTAAAGAGGGGGTCCGCTTATCAGACAG-CCGTGGCG 560
 QY 1130 aGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSe 1150
 Db 561 AGCATATCGCAAAAGGTTAGCCCTAGGCGACGGGATGTG-----CCTGCGTACAAGAA 614
 QY 1150 r 1150
 Db 615 C 615

RESULT 3
 AAH77704
 ID AAH77704 standard; cDNA; 3591 BP.
 XX AC AAH77704;
 XX DT 13-NOV-2001 (first entry)
 XX DE Nucleotide sequence of a plant SGS2 polypeptide.
 XX KW SGS2; RNA-dependent RNA polymerase; transgene silencing;
 XX KW transgene stability; crop plant; viral resistance; ss.
 XX OS Arabidopsis thaliana.
 XX PH Key Location/Qualifiers

| | | | |
|---|----------|--|------|
| FT | CDS | 1..3591 | |
| FT | | /*tag= a | |
| FT | | /product= "SGS2" | |
| XX | PN | FR2804128-Al. | |
| XX | PD | 27-JUL-2001. | |
| XX | PF | 26-JAN-2000; 2000FR-00001007. | |
| XX | PR | 26-JAN-2000; 2000FR-00001007. | |
| XX | PA | (RHOB-) RHOBIO. | |
| XX | PA | (INRG) INST NAT RECH AGRONOMIQUE. | |
| XX | PI | Beclin C, Elmayan T, Mourrain P, Vaucheret H; | |
| XX | DR | WPI; 2001-543303/61. | |
| XX | DR | P-PSDB; AAG67235. | |
| PT | PT | A new plant SGS2 gene involved in encoding an RNA-dependent RNA | |
| PT | PT | polymerase and in transgene silencing, increases transgene stability and | |
| PT | PT | expression in transgenic plants when it is inactivated. | |
| XX | PS | Claim 4; Page 32-37; 46pp; French. | |
| XX | CC | The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA- | |
| CC | CC | dependent RNA polymerase and is involved in transgene silencing. | |
| CC | CC | Inactivation of SGS2 is used to increase transgene stability and | |
| CC | CC | expression in plants, particularly crop plants, especially maize, corn, | |
| CC | CC | barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants. | |
| CC | CC | Overexpression of SGS2 can be used to increase resistance to viral | |
| CC | CC | infection in plants | |
| XX | SQ | Sequence 3591 BP; 1022 A; 714 C; 858 G; 997 T; 0 U; 0 Other; | |
| Alignment Scores: | | | |
| Pred. No.: | 3 218-10 | Length: | 3591 |
| Score: | 328.00 | Matches: | 197 |
| Percent Similarity: | 37.17% | Conservative: | 126 |
| Best Local Similarity: | 22.67% | Mismatches: | 312 |
| Query Match: | 4.43% | Indels: | 234 |
| DB: | 4 | Gaps: | 42 |
| US-09-913-878A-2 (1-1402) x AAH77704 (1-3591) | | | |
| QY | 535 | PhePheGluLeuLeuLeuProSerProThrSerProSerProValProValVal | 554 |
| DB | 1396 | TTTGTGCTCCTATTGTGAAGGATTGACATCAAGTTCTTCTCCAGAACACCTACGTT | 1455 |
| QY | 555 | SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer | 574 |
| DB | 1456 | TTTAAAGAGTGAAGGCATA-----TTACCGATGGG-----TTTAAA | 1494 |
| QY | 575 | LeuValGlyArgGlnTrpArgAlaPheAlaLysAspAlaGlyTyArgLysProLeu | 594 |
| DB | 1495 | CTATGTGGTAGAAAATAC---AGTTTCTAGCATTCCTCAGCCAAAT----- | 1536 |
| QY | 595 | ArgGluPheGlnLeuArgAlaGluAspProLysProLysProLysGluArgValHisPhe | 614 |
| DB | 1537 | -----CAACTGAGA-----GACCGCTCTGCATGGTTC | 1563 |
| QY | 615 | PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValPro | 634 |
| DB | 1564 | TTTGCTGAACACGGG-----AAAACACGT----- | 1587 |
| QY | 635 | AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu | 654 |
| DB | 1588 | -----GTGTCAGATATAAAACATGGATG | 1611 |
| QY | 655 | LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln | 673 |
| DB | 1612 | GGGAAGTTCAAGACAGAAATGTG-----GCAAAATGTCTGCTAGGATGGC | 1659 |

| | | | |
|----|------|--|------|
| QY | 674 | LeuGlyLeuSerLysThrTyralaIleMetThrLeuGluProHisGlnIleArgHisHis | 693 |
| DB | 1660 | CTGTGTTTCTCTCCACATATGCCATGTAGATGTCATGCTCCTCAGAGTT----- | 1710 |
| QY | 694 | LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg | 713 |
| DB | 1711 | GACACTGAGGTTCCAGATATTGAGAGAAATGGTATGTTTCTCTCAGCAATTTGTACA | 1770 |
| QY | 714 | MetSerArgSerValAlaLysArgileArgAspValLeuGlyLeuGlyAspVal----- | 731 |
| DB | 1771 | ATCACACCTGACCTCGCTGACGAAGTAATGGAGAACTTAAGTTG---GATGTGCACTAC | 1827 |
| QY | 732 | ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal | 750 |
| DB | 1828 | AGCCCTTGCTTATCAGATACGTTACGCAGAGTTTCAAAAGGGTT----- | 1872 |
| QY | 751 | AspAspThrGlyAspGluAspTrpIleGluThrTyProSerGln----- | 765 |
| DB | 1873 | -----GTTGCTGTGGCCATCAAAAAGTGAATGAATCAGG | 1908 |
| QY | 766 | -----ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal | 782 |
| DB | 1909 | CTAGCCCTTCGACACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC | 1965 |
| QY | 783 | ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuLeuLeuProValLeu | 802 |
| DB | 1966 | TGTTCC---TGGACGAGGTTTCAACCTGGTCTTAAATCGGCAGATAATTACCTCTTA | 2022 |
| QY | 803 | GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuLeuAsn | 822 |
| DB | 2023 | TCC-----GTACTAGGTGTTCCGATGAATATTTCTGG | 2055 |
| QY | 823 | AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg | 842 |
| DB | 2056 | GATATGCAG-----GAATCCATGCTCTATAAATCAACCGCATCTTGTATGATACA | 2106 |
| QY | 843 | GlnTrpValTyArgSerTyArgSerArgAlaThrArgValSerHisGlyArgValPro | 862 |
| DB | 2107 | GATGTGCAATTGAAGTTCTCAGCGCATCATGTGCT----- | 2142 |
| QY | 863 | PheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGly | 882 |
| DB | 2143 | -----GAACAGGGAACACTGCAGCTATCATGTAGTCAGGT | 2181 |
| QY | 883 | PheAspProLysLysGlnLysTyLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys | 902 |
| DB | 2182 | TTCAAAACCAAAACCAGCGCATCTACGCGGATGTTGTTCTTCAGTCAAGATTGCACAA | 2241 |
| QY | 903 | CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrlerTyMet | 922 |
| DB | 2242 | CTCTGGGCTCTCAGAGAAAATCTCGTATTTTGTACTTCAGAAAGGTGGCTAATGGGT | 2301 |
| QY | 923 | IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer----- | 940 |
| DB | 2302 | TGCCTAGACGACGAGGATCTTGAACATGGCCAATGTTTATTCAAGTCTCTAAACCG | 2361 |
| QY | 941 | -----SerLysPheArgAspGluGluGluSerPhe | 950 |
| DB | 2362 | TCATAGAAAATTGTTCTCCAAACATGGGTCTCGTTTAAAGGAGACAAAGACAGATCTG | 2421 |
| QY | 951 | ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp | 970 |
| DB | 2422 | GAAGTAGTTAAAGGCTAT---GTAGCCATTGCTTAAGAATCCTTGTCTTCCACCGGGAT | 2478 |
| QY | 971 | IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle | 990 |
| DB | 2479 | GTAAGGATTTAGACGCTTTGATGTACCCACGCTGCATGCATGATGATGCTGCTTATT | 2538 |
| QY | 991 | PheSerThrLysGlyAspValProLeuAlaLysLeuSerGlyGlyAspTyArgGly | 1010 |
| DB | 2539 | TTCCCTCAGAAAGGTATAGGCGGCATACAACAAAGAGCTTCTGCGATGACCTTCGCGG | 2598 |

Db 3484 GCAGCTGATTCTTGCACGAATCAA 3510
|||||
RESULT 4
AAC63740
ID AAC63740 standard; cDNA; 3901 BP.
XX
AC AAC63740;
XX
DT 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 7.
XX
KW Maize; plant; RNA-directed RNA polymerase; FIS; gene mapping;
KW gene marker; plant virus resistance; plant breeding; es.
XX
OS Zea mays.
XX
PN W0200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-USC09105.
XX
PR 07-APR-1999; 99US-0128094P.
XX
(DUPO) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Wang Z, Sakai H, Cahoon RE, Wang J;
FI Zhong G;
XX
WPI; 2000-679376/66.
DR P-PSDB; AAB28531.
XX
X New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance.
XX
PS Claim 2; Page 40-41; 62pp; English.
XX
CC The present sequence is one of a number of cDNA molecules which encode
CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC probes for genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes
XX
SQ Sequence 3901 BP; 928 A; 975 C; 1058 G; 935 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.: 1.15e-09 Length: 3901
Score: 319.50 Matches: 290
Percent Similarity: 33.81% Conservative: 185
Best Local Similarity: 20.64% Mismatches: 528
Query Match: 4.32% Indels: 405
DB: Gaps: 62

US-09-913-878A-2 (1-1402) x AAC63740 (1-3901)

Qy 92 ProLysAlaHisIalaAspProAspThrLeuProTrp-----SerLysGluPro 107
||| | | | | | | | | | | | | | | : : : : : : : :
Db 24 CCCAAACGGATTCGGCGCACCGACCAC---CCATGGGATCGCTCCGGGGCGGGCAGCCT 80
||| | | | | | | | | | | | | | | : : : : : : : :
Qy 108 PrxArgAlaLaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeuGluValLeu 127
||| | | | | | | | | | | | | | | : : : : : : : :
Db 81 CCTCCGGCGCGCGCGCGGGGACCTGGTGACCAAGCATAGCTTGGTGGATTG 140
||| | | | | | | | | | | | | | | : : : : : : : :
Qy 128 AsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArgThrLeuSer 147
||| | | | | | | | | | | | | | | : : : : : : : :
Db 141 -----ATGCCACCGTCAAGG-----CGCTCGATCTCG 167
||| | | | | | | | | | | | | | | : : : : : : : :

Db 1991 TGATGGGATACGACTTTCCCTGAGCCGAGCATGACCAAGTTTGAG----- 2036
Qy 774 physhisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLe 794
Db 2037 -TCTAACCATACTGTTAGAGGTGGTCTCG---TGGACAAGTTTCAGCCAGAGTCTT 2092
Qy 794 uasnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgL 814
Db 2093 AAATCGTCAGATTATACATTACTCTCC----- 2120
Qy 814 nAlaIleGlyAspArgLeuLeuAspLeuGlnArgGlnPheSerGluGlnLysHisAl 834
Db 2121 -----TC 2122
Qy 834 aleuasnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaTh 854
Db 2123 CTTGAATGTCGGGATGCTATCTTTGCTCAAAATGAGGAGCCATGTTTATCTAATCTCAA 2182
Qy 854 xArgVal---SerHisGlyArgValProPhe-----LeuAlaGlyLeuProAspSerGl 871
Db 2183 CAATATTGTCAGACTGTAGTTGCTTTTGACATTGTAAACCGCTCTTGCTGAGCA 2242
Qy 871 nGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLe 891
Db 2243 AGCAACCACTGCAGCACTGATGTTGAGTGTGGCATTTTCACCTGGAACGTAGCCACACT 2302
Qy 891 uGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAs 911
Db 2303 GAAGCAATGCTGTAGCTATAGTCTCCTCAGCTGCTAGTCTCTTTGGAGAGACAAAG 2362
Qy 911 nIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGl 931
Db 2363 GATTTTGTGCCAAGGAAAGTGTGTGATGGCTGCCTTGATGAACCTGGGATCTTGA 2422
Qy 931 uGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheTh 951
Db 2423 GCAAGGGCAGTCTTTATCCGGGCATCATCT-----CC 2455
Qy 951 rLeuLeuSerAspCys----- 956
Db 2456 ATCACTCAATAATGTCGTGTAAGTATGATCAAGATTGTCGTGAGCAACACAAATGC 2515
Qy 957 -----AspValLeuValAlaArgSerProAlaHisPheProSerAspIle 971
Db 2516 AGAGACCATCTGGGACTACTGTAATGCCAAGAAATCCATCCCTTCACAGGGGATGT 2575
Qy 971 eGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePh 991
Db 2576 CCGAATCTTGAGCTGTTGATGTGCTGAACTGCATCACCTTGTTGATTGCTGCTCTT 2635
Qy 991 eSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAs 1011
Db 2636 CCCCAAGAAAGGTGAGAGCCGCGCGAATGAAGCATCTGGAGTGTATCTGTATGGGGA 2695
Qy 1011 pMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPr 1031
Db 2696 TCTATACTCTGTAACATGGATGGATGAACCTTATACCCCTGTAAAGAGAGTTGGAACCC 2755
Qy 1031 oLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMe 1051
Db 2756 TATGGAC-----TACTCCCCAGCTGAAGCAAAACAACCTGCCACGGCGAGT 2800
Qy 1051 tAlaSerHisGlyThrGlySerAlaLysGluGlnThrThrTyrAspMetIleGlnLys 1071
Db 2801 ATCCCAACAT-----GATATTGTT----- 2819
Qy 1071 sSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluAr 1091
Db 2820 -GGTTTC-----TTCTTGAAGAACATGTTAAAT-----GAGAA 2851
Qy 1091 gLeuCysTyrIleAsnAsnSer-----ValSe 1100
Db 2852 ACTGGGTCCATTAAGCAATGCTCATCTTGTTCACGCTGATATGACGGAGTATGCGCAAT 2911

Qy 1100 rAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGl 1120
Db 2912 GGATGAGAAGTGTATTACGTTGGCAGAACTAGCAGCAACTGCTGTGAGTCTCCCAAGAC 2971
Qy 1120 nGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuGlyGlyAl 1140
Db 2972 AGCAAAAATTGTTGTCATGCCA-----GCATCCCTTCGCCAAATATATATCTTGACTT 3025
Qy 1140 aleuSerLeuProAspProMet---TyrLysSerAspSerTrpLeuGlyArg----- 1156
Db 3026 CATGGAAAGGAGGATGCTATCTCTATAGATCAGAGAAGATCCTTGGAAAGGCTTTATCG 3085
Qy 1156 ----- 1156
Db 3086 GTCAATCCAAAGACCTCCAGCATGATTGTTGTCAGAGAAACTTGCACATCTAACAA 3145
Qy 1157 -----GlyGluProThrHisIleLeuAspTyrLeuLysPheSer--- 1169
Db 3146 TCTGCTTATGATGACAGATATGAAAGTGTGCTGTCAGCTGATTTCTCTCGAGTCTGTG 3205
Qy 1170 -IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAl 1189
Db 3206 GCAGTCAAGTGTCTCATATGAAACAACTAACGCACTGCTCAACCAATATGGCGTGG 3265
Qy 1189 aLysAspThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTrTh 1208
Db 3266 CACTGAAGCAGAGCTTGTGACAGACATATATGG-----TCGCTTCCCACTAC----- 3314
Qy 1208 rPhePheLysGluLeuSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLys 1228
Db 3315 -----AGCAGCAGAAAGCGGGGACATAAAGGAGAGGTGAA 3352
Qy 1228 sAsnArgIleGlyGluValGluLysGlyArgLeuValLysAsnLysGluMetAr 1248
Db 3353 GAATGCATCTACTGCTCTTACAGAGGATT---AGGACATTTTCGAAAGCATTTGTGAC 3409
Qy 1248 gAspSerLysAspProTyrProValArgValAsnGlnValTyrGlu-----Ly 1264
Db 3410 AGATCAACACAGAGATCTCTGATGATGAGAAAAGTCGGTTTACGAGATGAAGCCCTCCGC 3469
Qy 1264 sTrpCysAlaIleThr-----ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSe 1282
Db 3470 TTGGTACCAGGTAACTTACCATCTGAATGGGTCCAGAGTCAAGGGAATGTTCAGATC 3529
Qy 1282 rLysValIleArgLeu---LeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTr 1301
Db 3530 TGACTGTGAGGACATGCCAGCAAGGCTTAGCTTTGC-ATGGATCGCGG-TTGAGCACCTG 3587
Qy 1301 pAlaLeuLeuArg 1305
Db 3588 GCACGGATTAAGA 3600
RESULT 5
AAC63741
ID AAC63741 standard; cDNA; 2816 BP.
AC AAC63741;
XX AAC63741;
XX 07-FEB-2001 (first entry)
DT
XX
DE
XX
XX Rice RNA-directed RNA polymerase F1S sequence, SEQ ID NO: 9.
XX Rice; plant; RNA-directed RNA polymerase; F1S; gene mapping; gene marker;
XX plant virus resistance; plant breeding; ss.
XX Oryza sativa.
OS
XX
XX WO2000060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US009105.

```

XX 07-APR-1999; 99US-0128094P.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
XX Zhong G;
XX WPI; 2000-679376/66.
XX P-PSDB; AAB28532.
XX
XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
XX for controlling gene expression and providing mechanisms to engineer
XX plant virus resistance.
XX Claim 2; Page 45-46; 62pp; English.
XX
XX The present sequence is one of a number of cDNA molecules which encode
XX plant RNA-directed RNA polymerase proteins. The sequences are useful as
XX probes for genetically and physically mapping genes, and as markers for
XX traits linked to those genes. They are useful for controlling gene
XX expression and provide mechanisms to engineer plant virus resistance.
XX They are also useful for plant breeding to develop lines with desired
XX phenotypes
XX
XX Sequence 2816 BP; 782 A; 572 C; 725 G; 736 T; 0 U; 1 Other;
XX
Alignment Scores:
Pred. No.: 1.1e-09 Length: 2816
Score: 317.00 Matches: 208
Percent Similarity: 37.54% Conservative: 164
Best Local Similarity: 20.99% Mismatches: 341
Query Match: 4.29% Indels: 278
DB: 3 Gaps: 49
XX
US-09-913-878A-2 (1-1402) x AAC63741 (1-2816)
QY 463 PheArgGlyLysProPhePro-----GluLysProProAsn 474
DB 31 TTCTGGGGGACAGTTCCTCCAGTTGATCGGTGGGAGGCTGAGAGGGCACTGAAT 90
QY 475 AspVal-----PheValThrAlaMetThrGlyAsnPheGluSerLysGly 489
DB 91 CGGGTGGCCAGGACCCCAACTTCTCTGCAGCAAGGTGGGGATGACCACGGGAGGTG 150
QY 490 SerAlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAlaPro 509
DB 151 CGGAGCGGTGGTATC-----ACGCCACCACAGGCT 180
QY 510 LeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArg 529
DB 181 TATTGTCTG-----CCTCCAGAAGTGGAGCGCTCAAAACCGTGTCTTCGGCAT 228
QY 530 PheGly-----ProAspArgPhePheGluLeuLeuLeuProSerPro----- 543
DB 229 TACCATGAGGTGGCTGACAGGTTTGTAGGTCATCTTTATGGACGGGTATGACAGGTG 288
QY 544 -----ThrSerThrSerProSerValProProValValSer--- 555
DB 289 CTGAACAACAATGTCTCACTCTTCACTGCACCAATGTCAAGACTTGATGTGGAAT 348
QY 556 -----LysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGln 572
DB 349 TTTTCCAGCAGACACACGGGTACAAAGCGTGTACAGGATGTTGTTGACGGGGTTTC 408
QY 573 HisSerLeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLys 592
DB 409 CAC-----ATGTGTGGAGAGAGTAC---TCATTTCTCGCATCTCATCGAAC- 453
QY 593 ProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgVal 612
DB 454 -----CAGTTAAG-----GACAACTCAGCT 474

```

```

QY 613 HisPhePheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerVal 632
DB 475 TGGTTCCTTTGGCCGAG----- 489
QY 633 ValProAlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAsp 652
DB 490 -----CACAAAGACACA-----ACGGTGGAAACAATTAGGAAG 522
QY 653 TrpLeuGlnLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPhe----- 669
DB 523 TGG-----ATGGACGGTTCA-----CAAGTAAGAATGTCGAGAGTGT 561
QY 670 SerArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGln 689
DB 562 GCTCGAATGGGCAGGCTCTCTCATCCACATATGCAACTGTGCACATGGCGCGGATGAG 621
QY 690 IleArgHisHisLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAsp 709
DB 622 GTTGATGAGAGTTTGCATGATGTTGTGCAT-----AATGAGTACATTTTCCCGAT 672
QY 710 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 729
DB 673 GGAATTTGGCAAAATTAACCCACAGATCTTCATTTGGAAAGTTGCCGAGAGGCTGCAGCTGACA 732
QY 730 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly-----Met 745
DB 733 GATAACCCGCCATCTGCTTATCAGATCAGGTTTGTGGCTTCAAGGGTCTCATAGCTGTC 792
QY 746 TrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln 765
DB 793 TGG-----CAAGGACATGGTGATGGGACACGGCTTTCTCTAGGCCAAGCATG 840
QY 766 ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerVal 785
DB 841 AGGAAGTTTGAGTCTTAACCATTTG-----GTGTTAGGGGTG 876
QY 786 AlaSer-----GluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGlu 803
DB 877 GTCCTCTGGCAAAAGTTCACGACGAGTCTTAATCGACACATATAATATTGCTATCC 936
QY 804 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 823
DB 937 TCA-----CTGAATGTCCACAGATTTCTATCTTTTGGCAA 969
QY 824 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 843
DB 970 ATGCAAGACACCATGCTTCTTAAC-----CTCAACAATATTCTATCAGACAGAT 1020
QY 844 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 863
DB 1021 GTTGTCTTTTGAGGTTTAAACAACCTTCATGTGCT----- 1053
QY 864 LeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGlyPhe 883
DB 1054 -----GATGATGAAACATCGACGATTCGCTATGCTAGTCTGCTTT 1095
QY 884 AspProLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 903
DB 1096 GAACCTTAGAAGTGAACACACACTTGAAGCAATGCTCTTGGCTATAAGGTCGCACAATG 1155
QY 904 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 923
DB 1156 CAGGATCTTTTGAAAGAACAGGATATTGTCCAAGGAGGAGGTGTGATGGGCTGT 1215
QY 924 AlaAspPheTrpGlyValLeuGluGlu----- 932
DB 1216 CTTGATGAGCTGGGGTCTTTCGACCAAGGGCAGTGTCTTTATTCGGGCAACAGTTCATCA 1275
QY 933 ---AsnGluValHisValGlyPheSerSerLysPheArgAspGluGluSerPheThr 951
DB 1276 TTGAATAGTATTATTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAAACACAGAG 1335

```

| | | | | | | | |
|----|------|--|------|----------|---|--|------|
| QY | 952 | LeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIle | 971 | Db | 2212 | GAGAGCATAGCATCGATCAATCGAG | 2262 |
| Db | 1336 | GTCAATTTGGGT---ACTGTGGTAATAGCAAGAAATCCCTGTCTTCATCCAGGGAGTGC | 1392 | QY | 1301 | TpAlaLeuLeuAlaArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe | 1320 |
| QY | 972 | GlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePhe | 991 | Db | 2263 | CTGTACGAGATGAAGCCCTCTCCCTGGTACCAGGTGACCTATCAC | 2316 |
| Db | 1393 | CGCATCTTGAAGCAGTTGATGTGCCGCACTGATCATCTCTGGTTGATTTGGTGTTC | 1452 | QY | 1321 | ValTppGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro | 1340 |
| QY | 992 | SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp | 1011 | Db | 2317 | GTGGAGAATCTCAGGGGC | 2349 |
| Db | 1453 | CCCCAGAAAGGTGAGAGCCCATGCTAACGAGCACTCTGGAGCGATCTTGATGGGAT | 1512 | QY | 1341 | GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys | 1360 |
| QY | 1012 | MetAlaTppValCysTppAspProGluLeuValaspGlyPheValasnAlaGluMetPro | 1031 | Db | 2350 | GGTGGAAATACCGGCAAGCCTTAGC | 2397 |
| Db | 1513 | CTCTACTTTGTGACATGGGATGAGAACTTATACCTCCAGGCAAGAGAGCTGGAACCT | 1572 | QY | 1361 | LysPheThrLysGlnTyrValAlaArgLeuGlu | 1371 |
| QY | 1032 | LeuGlu-----ProAspLeuSerArgTyrLysLysLysAspLysThrThrPheLys | 1048 | Db | 2398 | -----TACTTGGCGGATAAAG | 2415 |
| Db | 1573 | ATGGACTACTCCCACTGAGCA-----AAA | 1599 | RESULT 6 | | | |
| QY | 1049 | GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluInThrThrTyrAspMet | 1068 | ID | AAS17845 | standard; cDNA; 3731 BP. | |
| Db | 1600 | CAACTT-----CGCGCCCAAGTATCTCAACATGATATC | 1632 | AC | AAS17845; | | |
| QY | 1069 | IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr | 1088 | DT | 08-MAY-2002 | (first entry) | |
| Db | 1633 | ATTGATTTCTTTAAAGAACATGATAGTGAATCTTTGGTAGATCTGTAATGCTCAT | 1692 | DE | Tomato RNA-directed RNA polymerase (RdRP) cDNA from clone HF. | | |
| QY | 1089 | LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleLeuSer | 1108 | XX | Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription; | | |
| Db | 1693 | GTGTTCATCTGATCTTACGAGATGATGTCATGGATGAGAGTGTATTCACCTAGCT | 1752 | KW | Cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen; | | |
| QY | 1109 | SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer | 1128 | KW | transgenic plant; transgenic animal; cancer; viral infection; | | |
| Db | 1753 | GAGCTAGCAGCAACTCGCGTGGACTTCCCAAGACTGGC----- | 1791 | KW | immunoprecipitation; immunolocalisation; RdRP/HF; mutant; gene therapy; | | |
| QY | 1129 | TpAlaGlnLeuArgGluLeuGlyGlyAlaLeuSer-----LeuProAsp | 1145 | XX | Lycopersicon esculentum. | | |
| Db | 1792 | -----AACTTCGATATGCCACACACACCTTAAACCAAAAGTCTACCTGAC | 1839 | FH | Key | Location/Qualifiers | |
| QY | 1146 | ProMet-----TyrLysSerAspSerTrpLeuGlyArg----- | 1156 | FT | mutation | replace(31,T) | |
| Db | 1840 | TTCATGGGAAGAAAGATGGACAATCTTATAATCAGAAAGATCTTTGGAAGCTTAT | 1899 | FT | mutation | /*tag= b | |
| QY | 1157 | -----GlyGluProThrHisIleIleAspTyrLeuLysPhe | 1168 | FT | mutation | replace(32,G) | |
| Db | 1900 | CGTTCAATCCAAGAGCCCTCCATGGT-----GATGTGTGTCAAA | 1941 | FT | mutation | /*tag= c | |
| QY | 1169 | SerIleAlaArgPro-----AlaIleAspLysGluLeuGluAlaPheHis | 1183 | FT | mutation | /*tag= d | |
| Db | 1942 | GAGGTTTGCACTCCAAATGATCTGCTTATGACATAGATCTGGAAGTTCTCTGGTGCATCA | 2001 | FT | mutation | /*tag= e | |
| QY | 1184 | AsnAlaMetLysAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu | 1203 | FT | mutation | /*tag= f | |
| Db | 2002 | GATTTCTCGAGTCTGGCAATGCAATGCAATGTTCA-----TACGACCGCAGCTG | 2052 | FT | mutation | /*tag= g | |
| QY | 1204 | AlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu | 1223 | XX | US2001023067-A1. | product= "RdRP protein" | |
| Db | 2053 | AGTGCCTGCTCAGTCAGTAC-----AGGCTCCGCACTGAAGCTGAACCT | 2097 | PD | 20-SEP-2001. | /EC_number= "2.7.7.48" | |
| QY | 1224 | PheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLys | 1243 | PF | 08-FEB-2001; 2001US-00782874. | replace(2017,G) | |
| Db | 2098 | GTGACA-----GGGCACATAAG-----TTCCTTTGTTAAG | 2127 | PR | 05-MAR-1997; 97US-00811583. | /*tag= d | |
| QY | 1244 | Asn-----LysGluMetArgAspSerLysAspProTyrProValArgValAsnGlnVal | 1261 | PA | (WASS/) WASENEGGER M. | replace(2264,T) | |
| Db | 2128 | AACAGCAGCAAGAAAGCAAGCGACATAAAGGAC-----AGGCTGAAGACTGCT | 2175 | PA | (RIED/) RIEDEL L. | /*tag= e | |
| QY | 1262 | TyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSer-----GlyAlaAsnTyr | 1280 | PI | Wassenecker M, Riedel L, Schiebel W, Sanger HL; | replace(3597,A) | |
| Db | 2176 | TAC-----TCTGACTACGCAAGGATGTTCAAAAGTACCTTT | 2211 | XX | WPI; 2001-595798/67. | /*tag= f | |
| QY | 1281 | AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr | 1300 | XX | New nucleic acid molecule encoding a polypeptide having the enzymatic | replace(3599,A) | |
| | | | | PT | activity of RNA-directed RNA polymerase, for modulating gene expression | | |
| | | | | PT | and treating cancer and virus infection in human and animals. | | |


```
QY 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
Db .....
Db 2633 GACATGATCCGCCAAGCAAGTCCAGCGGATGGATATCTCCAGCACCAGCATACAG 2692
QY 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
Db .....
Db 2693 TTGGACCATGAT----- 2704
QY 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
Db .....
Db 2705 -----GTCAATTCAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAT 2752
QY 1080 AsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCysTyrIle 1095
Db .....
Db 2753 GACAGTTTGGGAATCATGCAAAATGCCATGTCGTATTTCGACAGACA-----GAA 2803
QY 1096 AsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db .....
Db 2804 CTGTATATGCCATGATGATCCATGCMAAAACTGCTGACCTCTTTCAATTCAGTG 2863
QY 1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgGlu 1135
Db .....
Db 2864 GACTTTCCAAAGACTGGTGTCCCGTGAAATACCATCT-----CACTTGGC----- 2911
QY 1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuG1 1155
Db .....
Db 2912 -----CCTAAGAAATACCCAGACTTCATGGATAG 2941
QY 1155 YargGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaAspProAlaI1 1175
Db .....
Db 2942 CCGG-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrG1 1193
QY 1175 eAspLysGlu----- 3001
Db 2960 ATCTCAGAAAGATTATTGGAAGCTTTTC-AGSAAAGTGAAG----- 3048
QY 1193 uAspGlyAlaHisPheTrpAspLeuAlaSerTyrTyrTrpPhePheLysGluI1 1213
Db .....
Db 3002 -GACAAAGCA-----CCTCAGGCTAGCTTATCGCGACCTTCCACAGAGATG 3048
QY 1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyG1 1233
Db .....
Db 3049 TGCAGGAGATCATATGATGCTGATATGGAAGTTGATGATTTGAAGATTACATTGACGA 3108
QY 1233 uVal-----GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
Db .....
Db 3109 AGCTTTTGACTACAAACTGATAT-----GACAACAAGCTGGGTAAATT 3153
QY 1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
Db .....
Db 3154 AATGGACTACTATGGCATAAAACAGAG-----GCTGAATACTTTAG 3195
QY 1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp-----Se 1282
Db .....
Db 3196 TGTGGCATATGAGGCTACAAACTTTGACCGCAGAAAGATGCTGAGGCCATTAG 3255
QY 1282 rLysValIleArgLeuLeuLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
Db .....
Db 3256 TGTGTGTGTGAGGCCTTGAG-----AAGGAGCAAGACCGCTGTT 3297
QY 1302 -----AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
Db .....
Db 3298 CAAGAGCGGTATGATATAGATGACATGTTACCAAGGCTTCGGTGTGTACCACTTAC 3357
QY 1313 rTyrHisLysSer-----ProLy 1319
Db .....
Db 3358 ATATCATCTACATATTTGGGTGTGTCACATCAGGGGTTGAAAGAGCTCAATTCATTAG 3417
QY 1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
Db .....
Db 3418 CTTTCCCTGTGTGTATGACAGCTTAATCCAGATTAAAGAGCAAGACGATTAACAG 3477
QY 1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
```

```
Db 3478 GCCAGTTCTCAACTTGTCTCAGGGCTCAACTGAGTCACAGATTAGTG----- 3529
QY 1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
Db .....
Db 3530 -----TTGAATGAGATTCCTCAGTCGAGCGGTTAAG 3558
QY 1379 pProGluValTyrGluValLeuGlyAspAsp 1389
Db .....
Db 3559 CTGATATATATATAATGTAATAGGTGTGAT 3589

RESULT 7
AAS17837
ID AAS17837 standard; cDNA; 3731 BP.
XX AC AAS17837;
XX 08-MAY-2002 (first entry)
XX Tomato RNA-directed RNA polymerase (RdRP) cDNA.
XX Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
XX cytosolic; virucide; RNA synthesis inhibitor; antibody; immunogen;
XX transgenic plant; transgenic animal; cancer; viral infection;
XX immunoprecipitation; immunolocalisation; ss; gene therapy.
XX Lycopersicon esculentum.
XX Key Location/Qualifiers
XX CDS 194..3538
XX /tag= a
XX /product= "RdRP protein"
XX /EC_number= "2.7.7.48"
XX US2001023067-A1.
XX 20-SEP-2001.
XX 08-FEB-2001; 2001US-00782874.
XX 05-MAR-1997; 97US-00811583.
XX (WASS/) WASSENEGGER M.
XX (RIED/) RIEDEL L.
XX Wassenegger M, Riedel L, Schtebel W, Sanger HL;
XX WPI; 2001-595798/67.
XX P-PSDB; AAU10006.
XX New nucleic acid molecule encoding a polypeptide having the enzymatic
XX activity of RNA-directed RNA polymerase, for modulating gene expression
XX and treating cancer and virus infection in human and animals.
XX Claim 1; Page 15-19; 34pp; English.
XX This sequence represents a cDNA encoding the tomato RNA-directed RNA
XX polymerase (RdRP) protein of the invention. The invention comprises the
XX nucleic acid and protein sequences of RdRP. The protein of the invention
XX can catalyze in vitro transcription of short single stranded RNAs into
XX DNA molecules, this transcription can be either primed by RNA or DNA
XX oligonucleotides or be unprimed. The protein may have cytostatic or
XX virucide activities. The sequences of the invention may be used in gene
XX therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
XX sequence and a template nucleic acid molecule derived from a nucleic acid
XX molecule which causes a disease are useful for treating a disease caused
XX by the undesired expression or overexpression of a nucleic acid molecule
XX in a human, rat or mouse, by administering the molecules. This system can
XX be used in the preparation of a pharmaceutical composition and for
XX inhibiting expression of any desired gene by transferring the RdRP system
XX to organisms that either lack a comparable mechanism or do not
XX sufficiently express their own RdRP. An antibody or an antagonist or
XX inhibitor to the protein are useful for inhibiting RNA directed RNA
```

CC synthesis and for ensuring stable heterologous, gene expression in transgenic organisms. The sequence is useful for probes and/or for the control of gene expression, as primers for amplification of nucleic acid molecules and as tools for the detection of expression of the cDNA molecules. Additionally, nucleotide and protein sequences are useful for suppression of undesired gene expression in humans and animals. The RARP is useful as a therapeutic agent for the control of cancer and virus infection in humans and animals and the antibody is useful for immunoprecipitation or immunolocalisation of the protein. Identification of polypeptides interacting with it and screening expression libraries

SQ Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other;

Alignment Scores:

| Pred. No.: | 2,3e-08 | Length: | 3731 |
|------------------------|---------|---------------|------|
| Score: | 297.50 | Matches: | 189 |
| Percent Similarity: | 35.78% | Conservative: | 137 |
| Best Local Similarity: | 20.75% | Mismatches: | 339 |
| Query Match: | 4.02% | Indels: | 247 |
| DB: | 4 | Gaps: | 35 |

US-09-913-878A-2 (1-1402) x AAS17837 (1-3731)

QY 562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
 DB 1349 GAAGACATAGATAACTTTCTCGTGT-----TCTTTTGTGTGATGAGGTGGAG 1399

QY 582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
 DB 1400 AAACGTATTCTACAGACTTA----- 1420

QY 602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
 DB 1421 ---TTACCAAAA-----GCAAGTACTGGAAGTGGT 1447

QY 622 PheArgProAspValPheLys-----ThrArgSerValValProAla 635
 DB 1448 CTCAGACAAACATCTATGAGAGATCTTCACTCTCGGAAAGGCTTTGTAATGGT 1507

QY 636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
 DB 1508 GATAAAAAATTGAA---TTTCTTGCAATTTTCATCGAGCCAGTTG----- 1549

QY 656 GlnLeuAspAsnThrTrp-----GlnProHisLeuLysLeu----- 668
 DB 1550 ---CGGATAATTTCAGTGTGATGTTGTCATCAAGACCTGCGCTTACTGCAAAATGATATA 1606

QY 669 -----PheSerArgile-----GlnLeu 674
 DB 1607 AGAGCTTGGATGGTGAATTTTCGAGATCAAGAAATGTCGCAAAATATGTCGCCAGACTT 1666

QY 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLysLys 694
 DB 1667 GGTCAATCTTTTGGTTCCTCCAGAGACACTTTG-----AGTGTCTTAGGCATGAG 1717

QY 695 ThrAspLeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
 DB 1718 ATTGAAGTATATCCGATGTAAAGTTTCATGGAACCCAGCTATGCTTTTCTGATGGAATT 1777

QY 712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
 DB 1778 GGTAAATATCTGTTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCCTCAATATACC 1837

QY 732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
 DB 1838 CCACTGCTTCCAGATTCGTTATGTTGGATATAAAGTGTGTGGGTGTGTATCCGGAT 1897

QY 752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
 DB 1898 TCATCAATGAAG-----TTGCTCTTGAGAAGAGCATGTCGMAATATGAATCAGAC 1948

QY 772 PheValaPheLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
 DB --- 1402

Db 1949 ACATA-----AAGTTAGATGTCCTTGG-----TGGAGCAAAATATCAGCCT 1990

QY 792 AlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgAlaArgAspLysValLys 811
 Db 1991 TGTATCTTAAATCGTCAACTGATTACGCTCTTGCT----- 2026

QY 812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
 Db 2027 ---ACACTTGGAGTGAAGATGAAGTTCTCGAACAGAAAGAAAGAACTGTAGATCAG 2083

QY 832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
 Db 2084 CTTGATGCTATCTTGTCATGATTCTTTGAAGGCACAGAGGCTTTGGAATTGATGCT--- 2140

QY 852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
 Db 2141 -----CCTCGAGAGAAC 2152

QY 872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
 Db 2153 ACTAATATTCTCAAGCAGATGCTAACTGTGTATAGCCTGATGCTGAGCCCTTCTT 2212

QY 892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
 Db 2213 TCAATGATGTTGCCAAACCTTCCGCGCATCCAAAGTTGCTCGATTTCGGGACTAGATCAAGA 2272

QY 912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
 Db 2273 ATATTATTCCAAATGAAGAAACAATGATGGGATGTTGGATGAATCCAGAACCTTGGAA 2332

QY 932 GluAsnGlnValHisValGlyPheSer-----SerLysPheArgAspGluGlu 947
 Db 2333 TATGTCAGTGTCTTGTTCAGTTTACTGTGTGTCGATGAGAGTGTTCACGATTTA 2392

QY 948 GluSerPhe-----ThrLeuLeuSerAspCysAspValLeu 959
 Db 2393 CATCCATTAAACAGCAGATCCCAACAGATTAATTCATCTCGAAGGAATATGGTT 2452

QY 960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
 Db 2453 GTTGCACAAAAATCCATGCTGTCATCCTGTGTATATTCTGTGTTTAAAGCTCTAAATGTT 2512

QY 980 ProGluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeu 999
 Db 2513 CGAGCGCTGCACCAATGATGTTGTTGTTTCCCTCAGAAAGGAAAGAAAGACTCAT 2572

QY 1000 AlaLysLysLeuSerGlyLysAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
 Db 2573 CGAATGAATGTTCTGGGAGTGTATGGATGGGATATCTACTTGTGTTGCTGGATCAA 2632

QY 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
 Db 2633 GACATGATCCCGCCCAAGCAAGTCCAGCGCATGGAATATCCTCCAGACCCAGCATAACAG 2692

QY 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
 Db 2693 TTGGACCATGAT----- 2704

QY 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
 Db 2705 -----GTCACAATTTGAGGAAGTTGAAGAGTACTTCCCACTATATATGTAAT 2752

QY 1080 AsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCysTyrIle 1095
 Db 2753 GACAGTTTGGGAATCATAGCAATGCCATGTCGATTATTGACAGACA-----GAA 2803

QY 1096 AsnAsnSerValSerAsnLysProAlaIleLeuLeuSerSerLeuValGlyAsnLeuVal 1115
 Db 2804 CCGTATATGGCCATGATGATGATCCATGCAAAACAACTTGTGAGCTCTTTTCAATTGCAGTG 2863

QY 1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
 Db 2864 GACTTTCACAGACTGTGTTCCCGCTGAAATACCATCT-----CAGTTGGCC----- 2911

| | | | |
|-----------|---|--|------|
| QY | 1136 | LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuG1 | 1155 |
| DB | 2912 | -----CCTAAGAAATACCAGACTTCATGGATAAG | 2941 |
| QY | 1155 | YargGlyGluProThrHisIleLeuAspTyrLeuLysPheSerIleAlaAagProAla11 | 1175 |
| DB | 2942 | -----ACAGACACGCTAT | 2959 |
| QY | 1175 | eAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrG1 | 1193 |
| DB | 2960 | ATCTCAGAAAGAGTATTGGAAAGCTTTTC-AGGAAAGTGAAG----- | 3001 |
| QY | 1193 | uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGlu11 | 1213 |
| DB | 3002 | -GACAAAGCA-----CCTCAGGCTAGCTCTATCGCGACCTTCACAAAGAGATGT | 3048 |
| QY | 1213 | eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyG1 | 1233 |
| DB | 3049 | TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGATTTGAAGATTACATTGACGA | 3108 |
| QY | 1233 | uVal-----GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe | 1250 |
| DB | 3109 | AGCTTTTGACTACAAAGCTGAATAT-----GACAAAGCTGGGTAATTT | 3153 |
| QY | 1250 | rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr | 1270 |
| DB | 3154 | AATGGACTACTATGGCATAAACACAG-----GCTGAATATTCTAG | 3195 |
| QY | 1270 | oGluAlaMetAspLysSerGlyAlaAsnTyrAsp-----Se | 1282 |
| DB | 3196 | TGCTGGCATTTATGAAGGCATCAAAACITTTGACCGCAGAAAGATGCTGAGGCCATTAG | 3255 |
| QY | 1282 | rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- | 1301 |
| DB | 3256 | TGTTGTGTGTGAGGCTCTGAG-----AAGGAGCAAGAGCTGTGT | 3297 |
| QY | 1302 | -----AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy | 1313 |
| DB | 3298 | CAAGAGCGCTAATGATATAGATGACATGTTACCAAGGCTTCGGCTTGTATCCAGCTTAC | 3357 |
| QY | 1313 | rTyrHisLysSer-----ProLy | 1319 |
| DB | 3358 | ATAATCTACATATTTGGGTTGCTACATCAGGGTTGAAAGAGCTCATTTCTATTAG | 3417 |
| QY | 1319 | spheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr | 1339 |
| DB | 3418 | CTTTCCCTGTGTGTATTATGACCAAGCTAATCCAGATTAAGAAGGACAAAGCAGTAACAG | 3477 |
| QY | 1339 | qProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs | 1359 |
| DB | 3478 | GCCAGTTCTCACTGTCTATCTCTCAGGCTCACTGAGTCAAGATTAGTG----- | 3529 |
| QY | 1359 | pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs | 1379 |
| DB | 3530 | -----TTGAAATGAGATTCCAGTCGAGCGGTTAAG | 3558 |
| QY | 1379 | pProGluValTyrGluValLeuGlyAspAsp | 1389 |
| DB | 3559 | CTGATATATATATATGTAATAGGGTGTGAT | 3589 |
| RESULT | 8 | | |
| AAAD04370 | | | |
| ID | AAAD04370 | standard; cDNA; 3731 BP. | |
| XX | XX | | |
| AC | AAAD04370; | | |
| XX | XX | | |
| DT | 04-JUL-2001 | (first entry) | |
| XX | XX | | |
| DE | Tomato RNA-directed RNA polymerase (RdRP) cDNA. | | |
| XX | Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression; | | |
| KW | transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss. | | |

| | | | |
|---|---|---|---|
| XX | | Lycopersicon esculentum. | |
| OS | XX | Key | Location/Qualifiers |
| PH | XX | CDS | 194..3538 |
| PT | XX | /tag= | a |
| PT | XX | /product= | "tomato C-protein having RNA-directed RNA polymerase (RdRP) activity" |
| XX | US6218142-B1. | | |
| FN | 17-APR-2001. | | |
| XX | XX | 05-MAR-1997; | 97US-00811583. |
| XX | XX | 05-MAR-1997; | 97US-00811583. |
| PR | (WASS/) WASSENEGGER M. | | |
| PA | (RIED/) RIEDEL L. | | |
| XX | Wassenegeger M, Riedel L, Schiebel W, Sanger HL; | | |
| XX | WPI; 2001-289830/30. | | |
| DR | P-PSDB; AAE00897. | | |
| XX | New nucleic acid molecules encoding polypeptides with RNA-directed RNA polymerase enzymatic activity, useful in modulating gene expression in plants, humans and animals, as well as in plant cell/tissue cultures or plant breeding. | | |
| XX | Claim 1; Col 25-34; 31pp; English. | | |
| PS | XX | The present sequence is a cDNA encoding tomato C-protein having RNA-directed RNA polymerase (RdRP) activity. This protein is capable of RNA-directed RNA synthesis, thus using RNA as a template for synthesising complementary RNA molecules. RdRP nucleic acid is useful for modulating gene expression in plants, humans and animals. This may lead to various physiological, developmental and/or morphological changes. Transgenic plants containing RdRP nucleic acid is especially useful in plant cell or tissue cultures and in plant breeding. RdRP is useful in gene therapy, particularly for treating a disease that is caused by the undesirable expression or overexpression of a gene | |
| XX | Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other; | | |
| Alignment Scores: | | | |
| Pred. No.: | 2,3e-08 | Length: | 3731 |
| Score: | 297.50 | Matches: | 189 |
| Percent Similarity: | 35.78% | Conservative: | 137 |
| Best Local Similarity: | 20.75% | Mismatches: | 339 |
| Query Match: | 4.02% | Indels: | 247 |
| DB: | 4 | Gaps: | 35 |
| US-09-913-878A-2 (1-1402) x AAD04370 (1-3731) | | | |
| QY | 562 | GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg | 581 |
| DB | 1349 | GAAGACATAGATAAATCTTCTTCGTGT-----TCITTTGTCATGAGGAGTGGAG | 1399 |
| QY | 582 | AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla | 601 |
| DB | 1400 | AAACTGTATTCTACAGACTTA----- | 1420 |
| QY | 602 | GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr | 621 |
| DB | 1421 | ---TTACAAA-----CCAACTACTGGAAGTGT | 1447 |
| QY | 622 | PheArgProAspValPheLys-----ThrArgSerValValProAla | 635 |
| DB | 1448 | GTGAGACAAACATCTATGAGAGGATCTTATCAACTCTGCGAAAGCTTTGTAAATGGT | 1507 |
| QY | 636 | GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu | 655 |

Db 1508 GATAAAAAATTGAA---TTTCTTGCAATTTTCATCGAGCCAGTTG----- 1549
Qy 656 GlnLeuAspAsnAsnThrTrp-----GlnProHisLeuLysLeu----- 668
Db 1550 ---CGGGATAATTCAAGTGTGATGTTTCATCAAGACCTGGCCCTACTGCAATGATATA 1606
Qy 669 -----PheSerArgile-----GlnLeu 674
Db 1607 AGAGCTTGGATGGGTGATTTTCGAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666
Qy 675 GlyLeuSerLysThrTyAlaLeuMetThrLeuGlnProHisGlnIleArgHisLys 694
Db 1667 GGTCAATCTTTGGTTCCTCCAGAGAGACTTGG-----AGTGTCTTAGCGATGAG 1717
Qy 695 ThrAspLeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db 1718 ATTGAAGTTATTCGCGATGTAAGGTTCATGGAACCGACTATGCTTTCTGATGTAATT 1777
Qy 712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyAspVal 731
Db 1778 GGTAAATATCTGGTGACTTTGCTCATAGATTGCTCAAAATGTGGCTTTCAATATACC 1837
Qy 732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
Db 1838 CCAATCTGCTTCCAGATTGCTATGCTGGATATAAAGGTGTGTGGGTGTGATCCGGAT 1897
Qy 752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
Db 1898 TCATCAATGAAG-----TTGTCTTTGAGAAGAGCATGTCGAAATATGATATCAGAC 1948
Qy 772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db 1949 AACATA-----AAGTTAGATGCTCTTGA-----TGAGCAATATACAGCT 1990
Qy 792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db 1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT----- 2026
Qy 812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db 2027 ---ACACTTGGAGTGAAGATGAAGTCTTCGAAACAGAGCAAAAGAGAGCTGTATGATCAG 2083
Qy 832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyArgLysTrpSerSer 851
Db 2084 CTGTGATGCTATCTTCATGATCTTTGTAAGGACAGAGAGGCTTTGGAATTGATGCTCT--- 2140
Qy 852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db 2141 -----CCTGGAGAGAAC 2152
Qy 872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyLeu 891
Db 2153 ACTAATATCTCAAGCAATGCTAAACTGTGTGTTAAGCTGTATGCTGAGCCCTTTCTT 2212
Qy 892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db 2213 TCAATGATGTGCAAAACCTTCGCGCATCCCAAGTTGCTCGATTTTGGGACTAGATCAAGA 2272
Qy 912 IleAspValGlyArgSerAlaTyIleTyMetIleAlaAspPheTrpGlyValLeuGlu 931
Db 2273 ATATTATTCNAATGGAAGACATGATGGATGTTTGGATGATCCAGAACCTTTGGAA 2332
Qy 932 GluAsnGluValHisValGlyPheSer-----SerLysPheArgAspGluGlu 947
Db 2333 TATGTCACAGGTGTTGTTTCACTTACTGGTGTGTCGACATGGAGAGTTTCTTGACGATTA 2392
Qy 948 GluSerPhe-----ThrLeuLeuSerAspCysAspValLeu 959
Db 2393 CATCCATTATAACAGCAGATCCACCACAGTAATTTCTATCTGAAGGGNAATGTGGTT 2452
Qy 960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
Db 2453 GTTGCAAAAAATCCATGCTGTGATATCTGTCATCTCTGATATCTGTTTAAAGGCTGTAATGTT 2512

Qy 980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
Db 2513 CGAGCCTGCACACATGATGATGTTGTATTCCTCCAGAAAGGAAAAAGACCTCAT 2572
Qy 1000 AlaLysLeuSerGlyGlyAspTyArgAspGlyAspMetAlaTrpValCysTrpAspPro 1019
Db 2573 CCAATGAATGTTCTGGAGTGTATGGATGGGATATCTACTTTGTTGCTGGGATCAA 2632
Qy 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
Db 2633 GACATGATCCCGCAAGCAAGTCAGCCGATGGAAATATCTCCAGCACCCAGCATAACAG 2692
Qy 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
Db 2693 TTGGACCATGAT----- 2704
Qy 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
Db 2705 -----GTCAATTTGGAAGTTGAAGAGTACTTCAACCACTATATTTGTGAAT 2752
Qy 1080 AsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCysTyIle 1095
Db 2753 GACAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTGCGACAGACA-----GAA 2803
Qy 1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db 2804 CTGTATGTCGATGATGATGATCCATGCAAAACCTGTGAGCTCTTTTCAATTTGCGATG 2863
Qy 1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
Db 2864 GACTTTCCAAAGACTGGTGTCTCCGCTGAAATACCATCT-----CAGTTGCGC----- 2911
Qy 1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGln 1155
Db 2912 -----CCTAAAGATATCCAGACTTCATGGATAG 2941
Qy 1155 YargGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIle 1175
Db 2942 CCGG-----ACAAGACCATAT 2959
Qy 1175 eAspLysGlu-----LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGln 1193
Db 2960 ATCTCAGAAAGAGTTATTGGAAGCTTTTC-AGGAAAGTGAAG----- 3001
Qy 1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIle 1213
Db 3002 -GACAAAGCA-----CCTCAGCTAGCTCTATCGGACCTTCACAGAGATGT 3048
Qy 1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGln 1233
Db 3049 TGCAAGGAGATCATATGATGATGATGGAAGTTGATGGATTGAAGATTACATTGACGA 3108
Qy 1233 uVal-----GluLysGluTyGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
Db 3109 AGCTTTTGTACTCAAAACTGAATAT-----GACAACAAGCTGGTAAATT 3153
Qy 1250 rLysAspProTyProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
Db 3154 AATGGACTACTATGGCATAAAACACAG-----GCTGAATATCTTAG-3195
Qy 1270 oGluAlaMetAspLysSerGlyAlaAsnTyArg-----Se 1282
Db 3196 TGTGTCGATTTAGAGGCATCAAAACTTTTACCGCAGAAAAGATGCTGAGGCCATTAG 3255
Qy 1282 rLysValIleArgLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp--- 1301
Db 3256 TGTGCTGTGGGCCCTTGAG-----AAGGAGGCAAGAGCCTGGT 3297
Qy 1302 -----AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
Db 3298 CAAGAGCGTAATGATATAGATGACATGTTACCAAGAGCTTCGGCTTGGTACCAGCTTAC 3357

```

QY 1313 rTyrHisLysSer-----ProLY 1319
Db 3358 ATATCATCTCATATTTGGGTGCTACATCAGGGTTGAAAGAGCTCATTTTCATTAG 3417
QY 1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
Db 3418 CTTTCCCTGGTGTGTTTATGACCACTAATCCAGATTAAAGAGGACAAAGCAGCTAACAG 3477
QY 1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
Db 3478 GCAGTTCTCACTTGTCTATCTCTCAGGGCTCACTGAGTCACAGATTAGTG----- 3529
QY 1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
Db 3530 -----TTGAATGAGATTCCAGTCGAGCGGTTAAG 3558
QY 1379 pProGluValTyrGluValLeuGlyAspAsp 1389
Db 3559 CTGATATATATATATATGATAGGTGTGAT 3589

RESULT 9
AAH77703
ID AAH77703 standard; DNA; 6863 BP.
XX
AC AAH77703;
XX
DT 13-NOV-2001 (first entry)
XX
DE Nucleotide sequence of the SGS2 gene.
XX
KW SGS2; RNA-dependent RNA polymerase; transgene silencing;
KM transgene stability; crop plant; viral resistance; ss.
XX
OS Arabidopsis thaliana.
XX
KE Key
FH Location/Qualifiers
FT promoter 1..850
FT exon /*tag= a
FT exon 851..3564
FT intron /*tag= b
FT intron 3565..3986
FT exon /*tag= c
FT exon 3987..4862
FT terminator /*tag= d
FT terminator 4863..6863
FT terminator /*tag= e
XX
PN FR2804128-A1.
XX
PD 27-JUL-2001.
XX
PF 26-JAN-2000; 2000FR-00001007.
XX
PR 26-JAN-2000; 2000FR-00001007.
XX
PA (RHOB-) RHOBIO.
PA (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI Beclin C, Elmayer T, Mourrain P, Vaucheret H;
XX
DR WPI; 2001-543303/61.
XX
PT A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT polymerase and in transgene silencing, increases transgene stability and
PT expression in transgenic plants when it is inactivated.
XX
PS Claim 2; Page 30-32; 46pp; French.
XX
CC The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-
CC dependent RNA polymerase and is involved in transgene silencing.

```

```

CC Inactivation of SGS2 is used to increase transgene stability and
CC expression in plants, particularly crop plants, especially maize, corn,
CC barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC Overexpression of SGS2 can be used to increase resistance to viral
CC infection in plants
XX
SQ Sequence 6863 BP; 1862 A; 1354 C; 1485 G; 2162 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.45e-07 Length: 6863
Score: 289.50 Matches: 190
Percent Similarity: 35.43% Conservative: 120
Best Local Similarity: 23.71% Mismatches: 297
Query Match: 3.91% Indels: 259
DB: 4 Gaps: 41
US-09-913-878A-2 (1-1402) x AAH77703 (1-6863)
QY 535 PhePheGluIleLeuIleProSerProThrSerThrSerProSerValProValVal 554
Db 2245 TTTGTGCTCCTATTGTGAAGGATTGTGATCATGTTCTTTCCAGAGACCTACGTT 2304
QY 555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
Db 2305 TTTAAAAGAGTGAAGAGCATA-----TTAACCGATGGG---TTTAAA 2343
QY 575 LeuValGlyArgGlnTrpArgAlaPheAlaLysAspAlaGlyTyrArgLysProLeu 594
Db 2344 CTATGTGGTAGAAATAC---AGTTTCTAGCATCTCAGCCAAAT----- 2385
QY 595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
Db 2386 -----CAACTGAGA-----GACCGCTCGCATGGTTC 2412
QY 615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
Db 2413 TTTGCTGAAGACGGG-----AAACACGT----- 2436
QY 635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
Db 2437 -----GTGTCAGATATATAAAACATGGATG 2460
QY 655 LeuGlnLeu---AspAenAenThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
Db 2461 GGGAAAGTTCAAAGACCAAGAAATGTG-----GCAAAATGTGCTGTAGATGGGC 2508
QY 674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
Db 2509 CTGTGCTTCTCTCCACATATGCCACTGTAGATGTCTGCTCAGAGGT----- 2559
QY 694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
Db 2560 GACACTGAGGTTCCAGATATTGAGAAATGGGTATGTTTCTCTCAGGAATTGGTACA 2619
QY 714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal----- 731
Db 2620 ATCACACTGACCTCGCTGACGAAGTAATGGAGAACTTAAGTTG---GATGTGCATAC 2676
QY 732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
Db 2677 AGCCCTTGTGCTTATCAGATACGTATTACGACGAGTTTCAAGGGGTT----- 2721
QY 751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln----- 765
Db 2722 -----GTGCTCGTTGGCCATCAAAAAGTGTGATGATCAGG 2757
QY 766 -----ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
Db 2758 CTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 2814
QY 783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeu 802
Db 2815 TGTTC---TGGACGAGGTTTCAACCTGGGTCTTTAAATCGGAGATATTACCTCTTA 2871

```

QY 803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
Db 2872 TCC-----GTAAGTGGTTCGGGATGAATAATCTCG 2904
QY 823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
Db 2905 GATATGCAG-----GAATCATGCTCTATAAAGTGAACCCATCTGTATGATGATACA 2955
QY 843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
Db 2956 GATGGGCATTTGAAGTTCTCAGGCATCATGTGCT----- 2991
QY 863 PheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsnSerGly 882
Db 2992 -----GAACAGGGAACACTGCAGCTATCATGCTTAGTGCAGGT 3030
QY 883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
Db 3031 TTCAACCAAAACCCAGCGCATCTACGCGGATGTTGCTTCTCAGTCAGAAATGACAA 3090
QY 903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
Db 3091 CTCTGGGCTCTCAGAGAAAATCTCGTATTTTGTACTTCAGGAAGTGGCTAATGGGT 3150
QY 923 IleAlaAspPheTrpGlyValLeuGluAsnGluValHisValGlyPheSer----- 940
Db 3151 TGCCTAGACGAAGCAGGATCTTGAACATGCCCAATCTTATTCAGTCTCTAAACCG 3210
QY 941 -----SerLysPheArgAspGluGluGluSerPhe 950
Db 3211 TCTATAGAAAATTTCTTCCAAACATGGGTCTCGTTTAAAGGACAAAGACAGATCTG 3270
QY 951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
Db 3271 GAAGTAGTAAAGGTAT---GTAGCCATTGCTAAGATCCTTGTCTCCACCGGGAT 3327
QY 971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 990
Db 3328 GTAAGATTTAGAAAGCTGTGATGATACCCAGCTGCATCATGATGATGATGCTTATT 3387
QY 991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGly 1010
Db 3388 TTCCCTCAGAAGGTGATAGCCGATACAAACGAAGCTTCTGCGAGTGACCTTGACGGG 3447
QY 1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
Db 3448 GACCTGTACTTGTGCTTGGATCAGAAACTCATC----- 3483
QY 1031 ProLeuGluProAspLeuSerArgTyr-----LeuLysLysAspLysThrThrPheLys 1048
Db 3484 -----CTCCCAACAGAGAAAGCTATCCGCCATGCAATTATGATGACGCTGAAGAG 3537
QY 1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr----- 1066
Db 3538 AGTTTA-----GGCGTGTCTCAACACCAAGTCAA-TATCTACTT 3578
QY 1067 -----AspMetIleGlnLysSerPhe 1073
Db 3579 CCCCCTTTTCTGTAGTGTAGATGATGACGCTTAAGATAAAATGGACAAGTGATTC 3638
QY 1074 HisPhe-----AlaLeuGlnProAsnPheLeuGlyMetCysThr 1086
Db 3639 CAGTATAACTCTCTGATATTTCTGAAGTGTGGAAACTTTGTTTGGATGCTGTAGT 3698
QY 1087 AsnTyr-----LysGluArgLeuCysTyr 1094
Db 3699 GATTACTTTTGTGCTTCTGAGTTCATAGATTGGTGGCATAGTGTGTTAAACAGCATAC 3758
QY 1095 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 1114
Db 3759 TGTGTTCTCTGTAGGAGCTAAG---ATGATGACATTTAAGAAACTTTGTCTATATT 3815

QY 1115 ValAspGlnSer-----Lys 1119
Db 3816 GTTGATCTGATTTGATTTTTCGATCGCTTACTGTGACTGTGACTGTTTGTACTAT 3875
QY 1120 GlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu----- 1135
Db 3876 CAAGGCACATCATGATGCTAGTACTGCTCCCAATGATGATGATTTCTTTGTTATT 3935
QY 1136 -----LeuLeuGlyGly-----AlaLeuSerLeuProAspProMet 1147
Db 3936 TTTCTGTGCTGTGCTTCCCAAGTCTTGGTAATGATGCTTTTGTACAGAC----- 3989
QY 1148 TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLys 1167
Db 3990 -----ATAATCGATTTC----- 4001
QY 1168 PheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 1187
Db 4002 -----TTTGCAGAAACTTGGCAATGACAGAGTTGGGCACAAATTTGCAAT----- 4046
QY 1188 AlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 1207
Db 4047 -----GCACACGTCGTTTCATGCTGATAGAGTGAAGTATGGA 4082
QY 1208 ThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeu 1227
Db 4083 GCCATGCGACGAAGAA-----TGTTTGTCTACTGGCAGAACTA 4118
QY 1228 LysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal----- 1242
Db 4119 GCTGCCACTGCGATGATTTCCTCAAGACA-----GGGAAATTTGTCTCATGTCCTCCAC 4175
QY 1243 ---LysAsnLysGluMetArgAsp-----SerLysAspProTyrProVal----- 1256
Db 4176 CTAAACCAAACTCTACCCAGATTTACCCGAAAGAAAGACTACCAAACTTACAAGTCG 4235
QY 1257 -----ArgValAsnGlnValTyrGluLysTrpCys----- 1266
Db 4236 AACAAATCTTGGTGGCTTTACACAGCGGTAAGAGGTTTATGATGAAGATCGAGAA 4295
QY 1267 AlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArg 1286
Db 4296 GCTTCTCTCAGAAGAAAGACACACCCAGTGCATCCCTTATGACGT----- 4343
QY 1287 LeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp 1301
Db 4344 GTTCTTGAATACCGGATTTGAAGATTTGATCCTCCGAGGCATGG 4388
RESULT 10
AAC63742
ID AAC63742 standard; cDNA; 3807 BP.
XX AAC63742;
XX
XX 07-FEB-2001 (first entry)
XX
XX Soybean RNA-directed RNA polymerase FIS contig, SEQ ID NO: 11.
XX Soybean; plant; RNA-directed RNA polymerase; FIS; gene mapping;
XX gene marker; plant virus resistance; plant breeding; ss.
XX Glycine max.
XX
XX WO200060097-A1.
XX
XX 12-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US009105.
XX
XX 07-APR-1999; 99US-0128094P.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX (FION-) PIONEER HI-BRED INT INC.

XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
 PI Zhong G;
 XX WPI: 2000-679376/66.
 DR P-PSDB; AAB28533.
 XX New nucleic acid fragments encoding RNA-directed RNA polymerase useful
 PT for controlling gene expression and providing mechanisms to engineer
 PT plant virus resistance.
 XX Claim 2; Page 49-50; 62pp; English.
 XX The present sequence is one of a number of cDNA molecules which encode
 CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
 CC probes for genetically and physically mapping genes, and as markers for
 CC traits linked to those genes. They are useful for controlling gene
 CC expression and provide mechanisms to engineer plant virus resistance.
 CC They are also useful for plant breeding to develop lines with desired
 CC phenotypes
 XX
 SQ Sequence 3807 BP; 1132 A; 662 C; 826 G; 1185 T; 0 U; 2 Other;
 Alignment Scores:
 Pred. No.: 1.76e-07 Length: 3807
 Score: 283.00 Matches: 234
 Percent Similarity: 32.43% Conservative: 160
 Best Local Similarity: 19.26% Mismatches: 402
 Query Match: 3.83% Indels: 420
 Ds: 3 Gaps: 53
 US-09-913-878A-2 (1-1402) x AAC63742 (1-3807)

QY 306 ProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuLeuGluSer 325
 DB 135 CCACAGCATCTAGAGCTTCGAT-----ATTTCGAAGAGGAAAGCGCAAGAGATTA 188
 QY 326 ProSerLys---GlyArgThrLysSerHisLeuAspAenGlnProLeuSerSer 344
 DB 189 GCATTAGGATGGGAAACAACTTGGATGGAT-----GGATTCCCT-ACA 233
 QY 345 SerGlnGlyGluThrSerPheSerThrTyrrGluSerPheProSerSerGly----- 362
 DB 234 TCTGTGAATGCTGTGATGAAGACATTTGAGACGATATCTGTGTGAAGGACACTGTG 293
 QY 363 -----GlyGluAlaLeuProGluPro----- 370
 DB 294 TTCGCCATTAAATTAAGACATGGAAGGTCGGTTCCAGAGCATTTGCCAATTATTCAA 353
 QY 371 -----SerArgSerAenGlyLeuAlaArgSer 379
 DB 354 TTCACACCGCMAATTCGTACATCTATGATGATCCAGAGTAAACATTTGAGACA 413
 QY 380 GluGluSerAlaArgSerGlnValGlnValHisAlaProValAlaAlaArgLeuArg 399
 DB 414 TTGCGGTATGGACCTCCTATTAAAGCTCGG-----GAAATGGAAGA 458
 QY 400 AsnIleTppProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGlu 419
 DB 459 GATATTGCCAAG----- 473
 QY 420 ValThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLys 439
 DB 474 ---CCAGGGGTGTTTCATAGTTGGATGATGAGAA-----CTGTCT 515
 QY 440 TyrAspProSerTrpSerThrAlaArgAspValThrAspIleTppLysThrLeuTyArg 459
 DB 516 TTGTGCTGTCAATCTCAAGGAAGA---TTCTCTGTTTATGGAAA----- 560
 QY 460 LeuAspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThr 479
 DB 561 -----AAGCAGGATGTTATTGTAAAT 581

QY 480 AlaMetThrGly-----AsnPheGluSerLysGlySerAlaValValLeuSer 495
 DB 582 TTGGGAGTGGAAATGAGAAAGATGCAATTTTATTTCCCAACAAGTGTGCAATACAAA 641
 QY 496 AlaValLeuAspTyr-----AsnProAspAsnSerPro 506
 DB 642 CTGTGACCTTCATATGAGACATTTGGAGATGAGCTGCATCGCCACCGAATGAGACT 701
 QY 507 ThrAlaProLeuTyLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeu 526
 DB 702 ACACGT-----TATCTG-----TTG 716
 QY 527 ThrArgArgPheGlyProAspArgPhePheGluLeuLeuLeuProSerProThrSerThr 546
 DB 717 ATTCAGTACTTGTGTCCTCCCGGGTGTGGAGACGATGA-----CTACATCAACA 770
 QY 547 SerProSerValProProValValSerLysGlnProGlyAlaValGluGluValleGln 566
 DB 771 AATATCTTTGATGATCCTTTGTTCAACTTCTTCAAAGATGCCCTGATGAG-----CAA 824
 QY 567 TrpLeu-----ThrMetGlyGlnHisSerLeuVal----- 576
 DB 825 TGATCCGAGCAATGATTTCACTCGAAGAGTGTGATTTGGCAGTCTCCGCCATATGT 884
 QY 577 -----GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArg 591
 DB 885 CTGGAGCTTCCTAATGGCGCACAACTTCCAAATTTTCAGGAAAACTTTGCTATTATGAG 944
 QY 592 LysProLeuArgGluPheGlnLeu----- 599
 DB 945 GAAAGTGAGGCAATACACTTTACAGACAGAGTTCCTTTCTCAAAATTTGGGGTCTT 1004
 QY 599 ----- 599
 DB 1005 GTCCCATTTGTCCTCTCTAGTGTGTTAAATATCATATGACATCTTTGTTAAAGTC 1064
 QY 600 -----Arg 600
 DB 1065 AATCATCTGGTTCAACATGATGCTTTGTCAGACCTGCATCTGATGATGATCTTATCGC 1124
 QY 601 AlaGluAspProLysProIleLeuLysGluArgValHisPhePheAlaGlu----- 617
 DB 1125 TTGGTTGATCCACGAGAGATGCCCGTGAATTTATTGAATATGCTTTAGAAAAGATTAC 1184
 QY 618 -----ThrGlyLeuPheArgPro-----AspValPheLysThr----- 629
 DB 1185 TATTCAAAGGAATTTTGTATGAACCCCAAGTGGCTGATCATCAGTACACAAAACATAC 1244
 QY 630 -----ArgSer----- 631
 DB 1245 CTGTAGTCAAAAATCATCTCGTCCCTGCAATATCCTTGGATACAGGGTTGTATATAC 1304
 QY 632 -----ValValProAlaGlu----- 636
 DB 1305 GTTCGAGGGTTTCAGATCAGCTTTCGAAAGPATATCTTTTGTGGTCCAGAGATGATGTC 1364
 QY 637 -----GluProValGluGln-----ArgThrGluPhe 645
 DB 1365 TCAATCTGTCTCCGTCATTTCCGTGAACATATTGATACTTTCTACGTTGTTCAITTT 1424
 QY 646 LysValSerGlnMetLeuAspTrpLeuGlnLeuLeuAsp-----Asn 659
 DB 1425 ---GTTGATGAAGAATTTGATAAAGTGTTCCTCACTGATTTATCATCAGTTTCACAGAAC 1481
 QY 660 AsnThrTrpGlnProHisLeuLysLeuPheSerArgLeuGlnLeuLeu----- 676
 DB 1482 AGAAAGTGAATATACACAGAAATTTTCCATCTCTTAAAGATGGCATAGTTGTTGTT 1541
 QY 677 SerLysThrTyAlaIleMetThrLeuGluProHisGlnIleArgHisLysThrAsp 696
 DB 1542 GATAAGAAAGTGTGAAATTTCTAGCATTTCTCATCAAGTCAAGTTCGCGGAAACATCTCTCTGG 1601
 QY 697 LeuLeuSerProSerGlyThrGly----- 704

Db 1602 ATGTTTCTCTTACAGAACTGGATGTACTGCTGCTTACATAAGGAAATGATGGGAAT 1661
Qy 704
Db 1662 TTAGCCAGATTAGGAATGTTGCTAAATATGCTGCTAGGCTGGGCAATCTTTGGTTCA 1721
Qy 704
Db 1722 TCTACTGAACTCTAAGTGTCCATAGGATGAAGTTGAATATATTCCTGTATGTGAAGAAG 1781
Qy 705
Db 1782 CTTACATATGATGGAACGAATATGCTTCTCTGATGGAATGGGAAATATCTTTGAA 1841
Qy 718 ValAlaLysArgIleArgAspValLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGly 737
Db 1842 TTGCCCCAGAAAGTGGCTAAAAAATGTGGTTATGATTGACCTCCATCTGCTTTAGATT 1901
Qy 738 ArgPheGlySerAlaLysGlyMetTrpValIleAspValAspThrGlyAspGluAsp 757
Db 1902 CGATATGTTGGGTACAAAGAGTTGTGCTGTTGACCTTAATCATGCTACAG----- 1955
Qy 758 TrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGln 777
Db 1956 ---TTATCACTGAGGAAGCATGCGGAATGATGATTCAGAT-----AAC 1997
Qy 778 ArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGln 797
Db 1998 ACAAGTTAGATGTTTGGCC---CGTAGTAGTTTCAGCCATGTTATCTGAATCGGCAG 2054
Qy 798 LeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGly 817
Db 2055 TTAATTTCTCTCTATCC-----ACTCTTGGTATCAAG 2087
Qy 818 AspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArg 837
Db 2088 GATGATGTTTGTGAGAAAAACAAAGAGAACTGTTAATCACTGAACATATATAACA 2147
Qy 838 ProValGluPheArgGlnTrpValTyrGluSerTyrSerArgAlaThrArgValSer 857
Db 2148 GATTCATTAAAGGCTCAGAAATTCGGACTTAATGCTGCT----- 2189
Qy 858 HisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPhe 877
Db 2190 ---GAGAGATC-----ACTAATGTTCTGAAGGAG 2216
Qy 878 LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 897
Db 2217 ATGCTCATTTGGTATACAGCTTAATCAAGAACCATTCCTTTCAATGATGCTTCAACA 2276
Qy 898 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 917
Db 2277 TTTAGGGCATCAAAACTTTTGGAAATGCGACTTAAATCTAGGATCTTTATTCCAAAAGGA 2336
Qy 918 AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisVal 937
Db 2337 AGAGCAATGATGGGATGCTAGATGAACCTAGAACCCCTAGAAATATGCTCAAGTATTTGTT 2396
Qy 938 GlyPheSerSer-----LysPheArgAspGluGluGluSerPheThrLeu--- 952
Db 2397 CAGTCTCTCTAACTAGGCTGAGAACTCTATCTCATGATATTTTTCATATGATTTGCCA 2456
Qy 953 -----LeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 969
Db 2457 AAGAAATTATATGTTAAAGGTAAAGTAGTTGTATGACAAAAYCCCTCTGTCGCCACCGGT 2516
Qy 970 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 989
Db 2517 GATGCGCTGTTTACAGCTGTGATGTCGACATTTGACACATGTTGAGGTGCTGTT 2576
Qy 990 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 1009

Db 2577 GTTTTCCCTCAAAAGGACCAAGACTCATCAAAATGAGTGTTCGGGAAGTATCTGGAT 2636
Qy 1010 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 1029
Db 2637 GGAGATATCTACTTCTGTTGTTGGGACCATGAATTGATT-----CCTTCT 2681
Qy 1030 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGln 1049
Db 2682 CGCCCAATTGATGCCA-----ATGGACTATATCTCTCCYCGCACT 2720
Qy 1050 LeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrSerMetIle 1069
Db 2721 GTGGAATTGGATCAT-----GATGTGATGATCGAGGAGTT 2756
Qy 1070 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr--- 1088
Db 2757 GAGGAGTATTTTGGCCAAATTACATGATCAATGACATGCTGGGAATAATTCCTCAATGCACAC 2816
Qy 1089 -----LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysPro 1103
Db 2817 ACTGTCTTTGCAGATAAGAACATTG-----AAAGCAATGCTGTGATCAA 2861
Qy 1104 AlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnIleVal 1123
Db 2862 TGTGTTAAGCTTGCAGAGTTGTTTCAACAGCAGTTGACTTTCTCTAAAACTGGTGTCCA 2921
Qy 1124 PheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeu-----LeuGlyGly 1139
Db 2922 -----GCTGTTATACCTCTGAACTTCATGTCACAAATATCCTGAC 2963
Qy 1140 AlaLeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeuGlyArg----- 1156
Db 2964 TTCATGGAGAGCTTACGACCATGAATGGAAGTTGATGGCTTCATGGATTATGTTGAT 3023
Qy 1157 -----GlyGluProThrHisIleIleAspTyrLeuLysPheSer 1169
Db 3024 AGGGAAGTGAAGAANAATATCAACAAGTCCGCTCAATTACATCTCTTCAAAAGTTG--- 3080
Qy 1170 IleAlaArgProAlaIleAspLysGluLeuGluAla----- 1181
Db 3081 GTTGCAGAGACTTTCAGACCATGAATGGAAGTTGATGGCTTCATGGATTATGTTGAT 3140
Qy 1182 -----PheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPhe 1198
Db 3141 GATGCTTTCTATCACAAAACCAATATGACTACAAAGTT-GGGAAATCTGATGGACTACTA 3199
Qy 1199 TrpAspProAspLeuAla-----SerTyrTyrThrPhePheLysGluIle 1213
Db 3200 TGGGATCAAAACTGAAGCTGAATCTCGGTGGGATATATGAAATGTCAAAATCTTT 3259
Qy 1214 SerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGly--- 1232
Db 3260 CAACAAAAGAGGAGTGCAGAAAGC-----AATCAATATGCTGT 3298
Qy 1233 GluValGluLysGluTyrGlyArg---LeuValLysAsnLysGlu 1246
Db 3299 GAGTCCCTTAAGGAAGAGGCGGCTGTTCAATGANAACAG 3343
RESULT 11
AAC63739
ID AAC63739 standard; cDNA; 3737 BP.
XX AAC63739;
AC AAC63739;
XX 07-FEB-2001 (first entry)
DT
DE Maize RNA-directed RNA polymerase EST sequence, SEQ ID NO: 5.
DE Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW Maize; gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX Zea mays.
XX

PN WO200060097-A1.
 XX 12-OCT-2000.
 PD 06-APR-2000; 2000WO-US009105.
 PF 07-APR-1999; 99US-0128094P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
 PI Zhong G;
 XX WPI; 2000-679376/66.
 DR F-PSDB; AAS28530.
 XX
 PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
 for controlling gene expression and providing mechanisms to engineer
 plant virus resistance.
 XX
 PS Claim 2; Page 35-36; 62pp; English.
 XX
 CC The present sequence is one of a number of cDNA molecules which encode
 plant RNA-directed RNA polymerase proteins. The sequences are useful as
 probes for genetically and physically mapping genes, and as markers for
 traits linked to those genes. They are useful for controlling gene
 expression and provide mechanisms to engineer plant virus resistance.
 CC They are also useful for plant breeding to develop lines with desired
 phenotypes
 CC
 XX
 SQ Sequence 3737 BP; 1063 A; 760 C; 863 G; 1051 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: Length: 3737
 Score: 267.50 Matches: 156
 Percent Similarity: 37.1% Conservative: 112
 Best Local Similarity: 21.64% Mismatches: 296
 Query Match: 3.62% Indels: 157
 DB: 3 Gaps: 27

US-09-913-878A-2 (1-1402) x AAC63739 (1-3737)

QY 645 PheLysValSerGlnMetLeuAsp-----TTP-----Leu 654
 DB 1472 TTTCTTCAAGTCAGCTTCGAGTAACTCTGCATGAGTGTTCCTCGGCGGGATTG 1531
 QY 655 LeuGlnLeuAspAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln--- 673
 DB 1532 ACTGCGAGCGACATAGGAAGTGGATGGGGACTTTCGAGATATCAGAAATGTGGCAAAG 1591
 QY 674 -----LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGln 689
 DB 1592 TATGCTGCAATATCTGGGCAATCTTCAGTCTCTCAACAGAACTTTAAAGTACACAAA 1651
 QY 690 IleArgHisHisLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAsp 709
 DB 1652 TCTGAGGTGGAACGAATCTCTGATATTACA---AATGGCACAAGTACATATTCTCTGAT 1708
 QY 710 GlyValClyArgMetSerArgSerValAla-----Lys 720
 DB 1709 GGAGTTGGAAAGATCTCAGCTAATTTGCAAGTGGAGGTGGCTATGAAGTGCAAAATTGAAA 1768
 QY 721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnIlyArgPheGly 740
 DB 1769 CGCTTT-----GCTCCTCTCTGTTTTTTCAGATAAGGTATGGC 1804
 QY 741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
 DB 1805 GGTTACAAAGGTGTTCTCGCTGTA-----GATACAGATCAATCAATGAAGTTCT 1855
 QY 761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780

DB 1856 TTGAGAAAAGCATGTCAAAGTTCAGTCAGAAAATAATC-----ACTCTT 1900
 QY 781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuPro 800
 DB 1901 GATGTCTCTGCATAC---AGCAAGTACCAACCATCTCTCTCGGAGTGGATTACT 1957
 QY 801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
 DB 1958 CTCTCTCTCA-----ACACTTGGGGTTAGCGATAATGTC 1990
 QY 821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
 DB 1991 TTTGAGCTAAAGCAGAAAGAGCCTTAAGGCAG-----TTGAACAGATGGTAAT 2041
 QY 841 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 860
 DB 2042 -----GAACACACAGGCTCTCGTGAACAGCTTGAACCTT 2074
 QY 861 ValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsn 880
 DB 2075 ATGCCCATG-----GGAGAGGTAAACCAATGTAGTTAAAGAAATTTGTGTCA 2119
 QY 881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
 DB 2120 TGTGGCTACAGCCTGATCATGAGCCCAATCTTTCATCTGCTGCTCAAACTTTAGAGCA 2179
 QY 901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
 DB 2180 TCCAGAGCTCTAGAGTTGAATAACAAAGTCAAGGATATTTCATCACAGCGGGGAGCAATG 2239
 QY 921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
 DB 2240 ATGGGTTGCTGGATGAACCTGCACACTTAAGTCGCGCAGGTATTCCTCCAGCTTCT 2299
 QY 941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
 DB 2300 TACAGTCAGATGACCATCCGCAAG-----GTGTTGTAACAGGAAAGTAGTTGTC 2350
 QY 961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
 DB 2351 GCCAAAATCCTTGTCTCCACCCTGTGACATACGCGGTCTCCAGCTCTTGTATGTTCT 2410
 QY 981 GluLeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAla 1000
 DB 2411 GCTCTGCACCACTGTTTGTACTGTGTGCTTCCACAGCAGGACCAAGGCGCACCT 2470
 QY 1001 LysLysLeuSerGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
 DB 2471 AATGAGTGTTCAGGGAGTATCTTGATGGGACATATATTTGTTCTTGGGATCCCAT 2530
 QY 1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
 DB 2531 CTTATTCCAAGTCGTTTGGTGGATCCTATGGACTACTCTCCAGCTTCAGCAGAAACATTA 2590
 QY 1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
 DB 2591 GACCATGAT----- 2599
 QY 1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
 DB 2600 -----GTCACTATTGGGAGATACAGGAGTACTTCACAACTACATAGTAAATGAG 2650
 QY 1081 PheLeuGlyMetCysThrAspTyr-----LysGluArgLeuCysTyr 1094
 DB 2651 AGTCTTGGGATTTACGCCAATGCGCATGTGGTGTCTTTACAGATCAGGAACGATG----- 2704
 QY 1095 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 1114
 DB 2705 -----AAAGCTGAGATCCACCGTGGTTCCTCACTGGCCCAAGCTCTCTCTAAGCT 2755
 QY 1115 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg 1134
 DB 2756 GTCGATTTCACAAAGACTGGATGCGG-----GCTCTGATTCACAT 2797

```

QY 1135 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet----- 1147
Db 2798 GAGCTA-----CATGTCAGAGATATCTGACTTCATGGAGAACTCGACAAGTCACC 2851
QY 1148 TyrLysSerAspSerTrpLeuGlyArg-----GlyGluProThrHisIle 1162
Db 2852 TATGAATCAAGGGTGTGATCGGGAAGCTCTATAGGGAATAAAGAACACACACAC 2911
QY 1163 IleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu-----Glu 1180
Db 2912 ATAAGCACTTCACAGGGAAGTGCAGGCGGCTTATGACACGATTTGATTGTCAT 2971
QY 1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
Db 2972 GCCTATGAAGATTACATTACTAGGCTATAGAGTTCAAGGAAGAGTAGCATTTTCAAGGCTG 3031
QY 1201 ProAspLeuAlaSerTyrTyrThrPheLysGlu-----IleSerAspLysSer 1217
Db 3032 GGTAATCTTATGACCACTATGGCATATAAAGTGAAGCTGACATAATAAGTGGATGATT 3091
QY 1218 ArgSerSerAlaLeuLeuPheThr-----ThrLysAsnArgIle 1231
Db 3092 CTAAGATGGCAAGAAATTCACCAAGTAGTGTGATGCTGATGCAATTAGATGGCGTG 3151
QY 1232 GlyGluValGluLysGlu-----TyrGlyArgLeuValLysAsnLysGluMet 1247
Db 3152 AGATCTTTGAGGAAGAAGCTAGTGTGTTCAATGAGATGACGACAGGAGAGGATGCG 3211
QY 1248 ArgAspSerLysAspPro-----TyrProValArgValAsnGlnValTyr 1262
Db 3212 CAAGATGCCATGGAGGCCAAGCCCTCTGCTGGTACCATTACTTATCATCAGCAGTAC 3271
QY 1263 -----GluLysTrpCysAla 1267
Db 3272 TGGGCGAGCTACATGAAGGGTATGATCGCGGCATCTTATTAGCTTCCCATGGTGTA 3331
QY 1268 IleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeu 1287
Db 3332 TATGACAAGTTGTGGCCATCAAGCAGGGAGGAATCTCCTCAGCAGATGGATCGAAAC 3391
QY 1288 Leu 1288
Db 3392 TTG 3394

RESULT 12
AAC63738
ID AAC63738 standard; cDNA; 1281 BP.
XX
AC AAC63738;
XX
DT 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase EST contig, SEQ ID NO: 3.
XX
KW Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS Zea mays.
XX
FN WO200060097-A1.
XX
PD 12-OCT-2000.
XX
PP 06-APR-2000; 2000WO-US009105.
XX
PP 07-APR-1999; 99US-0128094P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
PI Zhong G;

```

```

XX WFI; 2000-679376/66.
DR P-PSDB; AAB28529.
XX
PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance.
XX
PS Claim 2; Page 33; 62pp; English.
XX
CC The present sequence is one of a number of cDNA molecules which encode
CC plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC probes for genetically and physically mapping genes, and as markers for
CC traits linked to those genes. They are useful for controlling gene
CC expression and provide mechanisms to engineer plant virus resistance.
CC They are also useful for plant breeding to develop lines with desired
CC phenotypes
XX
SQ Sequence 1281 BP; 359 A; 282 C; 299 G; 338 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.: 3.21e-05 Length: 1281
Score: 236.00 Matches: 84
Percent Similarity: 40.10% Conservative: 74
Best Local Similarity: 21.32% Mismatches: 158
Query Match: 3.19% Indels: 78
DB: 3 Gaps: 11

US-09-913-878A-2 (1-1402) x AAC63738 (1-1281)
QY 698 LeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSer 717
Db 77 GTCACAACTGATGGCACTAAATACATATTTTCAGATGGTATCGGAGATTTTCTACTAGA 136
QY 718 ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal-----ProSerAlaVal 735
Db 137 TTGGCCAGACAGTCGCCAAATTAATTGGCTTAGACCCAGCTCATCTCTCTGCTTTT 196
QY 736 GlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAspThrGlyAsp 755
Db 197 CAATAAGTATGGGGGCTATAAGGA-----GTCATCACTATTGACCCCTCATCTCTT 250
QY 756 GluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLys 775
Db 251 TTCAAT---CTTCTCTGCGACCTAGTATGAAGAGTTTGAATCG----- 292
QY 776 HisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsn 795
Db 293 -----AAGAGCACTATGCTGAAC 310
QY 796 Leu-----GlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db 311 ATTACAAATTGGGTAAAGTCTCAGCATGTTATGTGAACCGTGAATTTATCTCTCTT 370
QY 812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db 371 TCAACATTGGGATAAAGGATGAAGTATTTGANTCGATGCAACAAGATGACATGCACGAA 430
QY 832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
Db 431 TCAGATGGAAATGCTAACAAATAAAGAA----- 457
QY 852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db 458 -----GCTGTTTGTCTCTCTAGGGAATAATGGTGGCGTGAT 496
QY 872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db 497 ACCAAGACGGCAGCTGATGCTTCTCAAGGCTATGAACCAAGTTTCAGAGCCCTTACCTA 556
QY 892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db 557 TTAATGATTTCTTAAGCCCATCGGCTAATAGCTGACCGACATAGAACTCGGTGTAAG 616

```

QY 912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTyrGlyValLeuGlu 931
 Db 617 ATTATGTCAGGAGCGCGTCTTCTTATGTTGGTGAACACTTGCAATTAGAG 676
 QY 932 GluAsnGluValHisValGlyPheSerSerIysPheArgAspGlu-----GluGlu 948
 Db 677 TATGCCCAAGTTTACATCAGATTACAAAGATCCCAAGGACGAAAGTACAGTGAACAG 736
 QY 949 SerPhe-----ThrLeuLeuSerAspCysAspValLeuValAla 961
 Db 737 CGGTCTTTTGCACGATGATGCGAAACAGCTGTAATGTTCGGAAGTTGCAATCACA 796
 QY 962 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheIysProGlu 981
 Db 797 AAAAACCCCTTGCTCCATCTCGTGTGATGTCAGAGTACTGAGCTGTATATGACCTGGA 856
 QY 982 Leu-----HisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
 Db 857 TTGGATGCTAGGGCTTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 916
 QY 1000 AlaLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTyrValCysTyrAspPro 1019
 Db 917 CGCAATGAATGTCGCGGGCGGATTTGGATGGCGGACCTCTCTTTTATTACTTGGGATGAC 976
 QY 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
 Db 977 AAACCTGATTCGGAGAGGTTGATGACCTATG----- 1009
 QY 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
 Db 1010 -----GACTACACTGCAACGAGCGCCACGCAATGACCATGCT----- 1048
 QY 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 1073
 Db 1049 -----GTTACACTTGAGGAATTCAGAGCACTTC 1078

RESULT 13
 ACA42620
 ID ACA42620 standard; DNA; 4551 BP.
 XX
 AC ACA42620;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #24277.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; Gene.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WC200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US0009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362899P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-023926/02.
 DR P-PSDB; ABU38750.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 14; SEQ ID NO 30490; 1765pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4551 BP; 779 A; 1608 C; 1519 G; 645 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.04 Length: 4551
 Score: 195.50 Matches: 271
 Percent Similarity: 32.95% Conservativeness: 189
 Best Local Similarity: 19.41% Mismatches: 477
 Query Match: 2.64% Indels: 465
 DB: 64 Gaps: 64

US-09-913-878A-2 (1-1402) x ACA42620 (1-4551)
 QY 86 AlaSerSerAsnTrp-ValProLysAlaHisAlaAspProAspThrLeuProTrpSerLys 105
 Db 257 TCGAACAGCGCTGGAGTCTCTCAAGCGCGCAAGAACTTGC-ACCGGTCCTGGCCGAG 315
 QY 105 sGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeuGl 125
 Db 316 GGACAGTTCAGGGCGGAGGCGCTGCGAACACCCCTGGCTATCTCT- 361
 QY 125 uValLeuAsnArgPheMetProProAsnAsnThrProGlyArgThrPheGlyArgTh 145
 Db 362 -----GGGACAACTGGTATCCCAAG 381
 QY 145 rLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAspGl 165
 Db 382 TCGCGCGGGTGATGCGAGCGCGCTTCTCTACAGTTCGCGGCAACAGGTCACCGAGCG 441
 QY 165 uProAlaAsnValThrPheAlaAspPro---ProLysArg-SerLeuThrArgSerAla 184
 Db 442 GTGCCCGAGCTGAAGATGGGCAACGCCATGACCGCGCCGACGCTACGACGGCGAGCG 501
 QY 184 hr-----GlyProProIleHisGlyAlaAlaIleProLeuLysPheProAsp 200
 Db 502 CAATACGCGCGGATTCGCGCTGCGGCTGCTGAGCCAGTTCGCTA-----TCCG 552

| | | | |
|----|------|--|------|
| Qy | 200 | roValAenThrGlySerLysArgProSerLeuGluSerGluAsn-----LeuAsnG | 217 |
| Db | 553 | CTGTGGAGCGCAAGCCATCGAGCGCGCGAGGACGTCGAAGCTGGCGGTTCCTCCAGC | 612 |
| Qy | 217 | lnCysThrLysArgAlaIyegLYLysLeuSerAspAsnValAlaAlaAlaAlaAlaProp | 237 |
| Db | 613 | GAGTTCATCAACGTCGTCCCGGCAATGCTCACTCCGCTGGGCAAGCTGTACACCGCGGC | 672 |
| Qy | 237 | roValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgA | 257 |
| Db | 673 | CTGTACCTGTGTCGAGGGCGCTTGTGGCAAGTACCG--CGCCACACCGTGGTTCGTC | 729 |
| Qy | 257 | spProThrAlaThrGlyHisArg--ArgAlaAspGlnValAspSerPheAspThrSerG | 276 |
| Db | 730 | TCCAACAGCTGGCGGTGAGCAGGTCCCGCGACGACGCTGCTGTGTGACCGCACGC | 789 |
| Qy | 276 | lnGlyThrSerTyrGlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrG | 296 |
| Db | 790 | AA-----GCACGAAGGCACCCCGTCCCGACACC | 815 |
| Qy | 296 | lnSerSerPheGluAlaProProSer-----GlnProArgGluLysArgP | 311 |
| Db | 820 | AAGTGTCTGACCGGACGGCTGGCGGTGATGAGCAGCGGCAATACGACGCGACGGC | 879 |
| Qy | 311 | roValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgT | 331 |
| Db | 880 | CTGCTGGCTCAAGCAGCGCCAGCTCGGAGCGCTCTACGT-----GATCGGCG-AGGA | 932 |
| Qy | 331 | hrThrLysSerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyClnThrSerP | 351 |
| Db | 933 | CCGCGAGG-----GGGGGTCTTCGCTCCGAGAACTTCTA | 968 |
| Qy | 351 | heSerThrTyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluPro- | 370 |
| Db | 969 | TTACGACAGCGAGATCTACGACACCAAGATCT-----ACGCCTTCACGA | 1013 |
| Qy | 371 | --SerArgSerAsnGlyLeuAlaArg-----SerGluGluSerAlaA | 384 |
| Db | 1014 | CCGCGCGGTGTACCGCGCGGCGAGCTGGTCTGTTGAAGATGTCGCGCGCGAGTTCAA | 1073 |
| Qy | 384 | rgSerGlnValGlnValHisAlaPro-----ValV | 394 |
| Db | 1074 | GGAGCCCGGCAATCCAGCGCGCGCGCGCGCGCTGAGCTGATCGACGC | 1133 |
| Qy | 394 | alAlaAlaArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProL | 414 |
| Db | 1134 | CAGCGGCACGGTATTGGCAAGCCCTGGACCTCC--GCTTCGACGCGAAGAGCGCGCAA | 1190 |
| Qy | 414 | euAlaValAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspG | 434 |
| Db | 1191 | TGGCGCGTTC-----AATCGCGGAG-AAGCGCTCGCGC | 1225 |
| Qy | 434 | luSerLeuGlyLeuLysTyrAsp-----ProSerTrpSerThrAlaArgAspValT | 451 |
| Db | 1226 | CGCGCTACGAGTTCGGCTTCGACTACCGCGGCGACAGCTACAGCAGCGCTTCGCGCTG | 1285 |
| Qy | 451 | hrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysPro----- | 467 |
| Db | 1286 | CCGAATACATCAAGCGGCACCTCGAAGTCGCGCTCGACCTGGCCAAAGCGGACTTCAAG | 1345 |
| Qy | 468 | -----PheProGluLysProPro----- | 473 |
| Db | 1346 | CCGCGGACCGGTGAAGGCGAGATGCTCTCTCTATCCGACGGCAAGCGCGGTGGCCA | 1405 |
| Qy | 474 | -----AsnAspValP | 477 |
| Db | 1406 | ACGCGCGCTCAACTGAGCTGCGCGCCAGCAGCTGTCTGATGGTGGACAACGAACGTG- | 1464 |
| Qy | 477 | heValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaValValleuSer---- | 495 |
| Db | 1465 | -----CAGTATCTCGGCAGTTCCGGTGGAACTGAGCAGCCCAACTGACCAACG | 1516 |
| Qy | 496 | -----AlaValLeuAspTyrAsnProAspAsnSerProThr----- | 507 |

| | | | |
|----|------|--|------|
| Db | 1517 | ACGCCAAGGGTCGCGCGCCCATGGAATTGCGCGCGCGAGAAACCCAGTCGCTCATGTC | 1576 |
| Qy | 508 | -----AlaProLeuTyrLeuValLysLeu-----LysProLeuMetP | 520 |
| Db | 1577 | TGACCATCTTCGCCAGCAGCGCGCCCTACCGGCTCAGACCAAGGATCCTCA | 1636 |
| Qy | 520 | heGluGInGlyCysArgLeuThrArgArgPheGlyProAspArgPhe | 535 |
| Db | 1637 | TCGAGCGCGCGCGCGCGCTACCGCTGAGCGCGCGCAGCGCTTCAGCGCGCGCGCG | 1696 |
| Qy | 536 | -----PheGluLeuLeuProSerProThrSerThrSerProSerValProP | 552 |
| Db | 1697 | AGAGGTCGAGTTGAGTATCCAGCGAAGCGCGACCCCGCTCAAGCCGAGC | 1749 |
| Qy | 552 | roValValSerLysGInProGlyAlaValGluGluValIleGInTrpLeuThrMetGlyG | 572 |
| Db | 1750 | -----AGCTACCAAGTGGATCCGCTCGAGG | 1774 |
| Qy | 572 | lnHisSer-----LeuValGlyArgGlnTrpArgAlaPhePheAlaLysA | 587 |
| Db | 1775 | ACCGCGCACCGCAGCGGTCGCGTCCGCGCGCGCTTCGCGCTGACCTTCGAGCGCC | 1834 |
| Qy | 587 | spAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProI | 607 |
| Db | 1835 | CCGGTACCTATTCGCTGAGCTGGC-----GACGACAAGGGCCAGC | 1876 |
| Qy | 607 | lelleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAsp | 625 |
| Db | 1877 | TCCTTGCGCTACCGGCCACTCGGTGAGCGCGAGGGCGTGAAGTCGGTCCGGGACCG | 1936 |
| Qy | 626 | -----ValPhe-----LysThrArgSerValValP | 634 |
| Db | 1937 | TCGAGGTGGTGTTCGACAAGCCCGAGTACCGCACCGGAGGAAGCCTCCGACTGATCA | 1996 |
| Qy | 634 | roAlaGluGluProValGlu-----GlnArgThrGluPheLysValS | 648 |
| Db | 1997 | CTTTCCCGCAGCGGTGGAGATCGCGTGTGTTCCTCGGAGCGCGACAAGGTTCGAGGCCA | 2056 |
| Qy | 648 | erGlnMetLeu-----AspTrpLeu-----LeuGlnLeuAspAsnAsnThr | 661 |
| Db | 2057 | CCGCCCTGTCTTCCAGGGCGCGACATGGCTGGCTGGCTGGAAAGCTCAATCCGACCCAGT | 2116 |
| Qy | 662 | -----TrpGlnProHisLeuLysLeuPheSer | 670 |
| Db | 2117 | ACCGCGTGTGATCCCGGTGCGGAGGAGTCTCTCGCGAACCTGACCTTCTCGGTGCTCT | 2176 |
| Qy | 671 | -----ArgIleGlnLeuGlyLeuSerLysT | 679 |
| Db | 2177 | ACACCAAGCGCGGAGCTACAGCTTCAGAACCCCGGGATCAAGGTTCGGCATCCGCCAGG | 2236 |
| Qy | 679 | hrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLysThr | 695 |
| Db | 2237 | TAGAGATCGACATCGCCACCGACGAGCGCTACGAACCGGGCGAGACGGTGACCGTGA | 2296 |
| Qy | 696 | -----AspLeuLeuSerProSerGlyThrGlyGluValMetA | 708 |
| Db | 2297 | CCCTGGCCACGCGCTTCGCGGCAAAACCGGTTCACAGCCACCTGACCGTCCGCGTGGTGG | 2356 |
| Qy | 708 | snAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyL | 728 |
| Db | 2357 | ATGAGATGGTCTATGCCCTGCAGCGGAGATCGCCC-CGGCATCGACCAAGTTCTT | 2411 |
| Qy | 728 | euGlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValI | 748 |
| Db | 2412 | -----CTACCATCCGCGCGCAACACGTGC-GCACCCAGCG- | 2446 |
| Qy | 748 | leAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--ArgL | 767 |
| Db | 2447 | -----CCAGCCTGGCTTCTATCAGCTACGA | 2471 |
| Qy | 767 | ysnTrpGluCys-----AspPheValAspLysHisGlnArgThrL | 780 |

Db 2472 COTGGCGTTGCCCGCAGACACAGCGCGACCCCGGACGG-GCGAACCGCAGCGCGCGCG 2530
 Qy 780 euGluVal-----ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAenLeuGlnL 798
 Db 2531 TCAGGTACTGAGCGCGCGCGCGCGGAGGACGTCACACCGCGCGCTGGCAGCGGAGC 2590
 Qy 798 euLeuProValLeuGluAspAgaAlaAgaAspLysValLysMet-----A 813
 Db 2591 TGGTCACCGACCCAGGAGGAGGCGAGCTTCAGCTCCGATGCCCGACTCGCTGACCC 2650
 Qy 813 tgGlnAlaileGlyAspArgLeuLeuLeuAsp-----LeuGlnArgGlnP 828
 Db 2651 GCTGGCGGATCACCGCGCGGCGCATCGACGACCAAGCGCGAGTCCGCGCAGAGCAGT 2710
 Qy 828 heSerGluGlnLysHisAlaLeu-----AsnArgProValGluPheArgGln- 843
 Db 2711 TCCITTCGTCGAGAGCGCGCTGTACCTGAAGTGGAGCGCGCGCTGCTTCGCGCAGG 2770
 Qy 844 -----TrpValTyrGluSerTyrSerSerArgAlaThrArgV 856
 Db 2771 GTGACCGACCGACCTCGCGCTGTTCGCTTC----- 2802
 Qy 856 alSerHisGlyArgValPro-----PheLeuAlaGlyLeuProAspSerGln- 871
 Db 2803 --AACCGAGCGGAAACAGCGCGGTCAAGCGCGAGCTCTCAGCGGCGCGCTGGCAGCAGC 2860
 Qy 872 --GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrL 891
 Db 2861 GCAGCCAGACCTGGAGCTGGCCAGGAGTCAACTACATTCGCTCGCGCAACAGCGCG 2920
 Qy 891 euGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuA 911
 Db 2921 TCAGCGATGGCGACTGGAGCGCGCGGATTCGCCAGGAGCGCGAGTCCGCGACCGCTGG 2980
 Qy 911 snIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuG 931
 Db 2981 CGTGGCTTCAC-----CTGCTGGCGGACGGCTGGGAGTGGAGC 3022
 Qy 931 luGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluSerPheT 951
 Db 3023 AG----- 3024
 Qy 951 hrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspI 971
 Db 3025 -----ATGCAGAACTCAGCTCGCGCGCGGCGAGCAACCCGCTGCAACTGCGCGGAGC 3079
 Qy 971 leGlnArgValArg-----AlaValPheLysProGluLeuH 983
 Db 3080 CGCGGACGTAGCGCTCGCGCTGGCGGATGGCGCGCGCGCTGCCTAC-----CTGG 3130
 Qy 983 isSerLeuLysAspValIleIlePheSerThrLysGlyAsp----- 996
 Db 3131 GCATCTCGATGATCTGTGGAGTATCCCTACGCGCGGCTCGAGCAGACCGCCAGCCAAC 3190
 Qy 997 --ValProLeuAla-----LysLysLeuSerGlyGlyAsp----- 1007
 Db 3191 TGCTGCGCTGAGCATCGCTATCCGCGCTGCGCGCGCGGCGAGCGCGGATTCGCGACC 3250
 Qy 1008 -----TyrAspGlyAspMetA 1013
 Db 3251 GCTGCGCTGATCATGACAGAACAGTCCGCTGCGCTGCGTGCAGATGCGCGCGCGGACG 3310
 Qy 1013 laTrpValCysTrpAspProGluIleValAspGly-----PheValAsnAla----- 1028
 Db 3311 CCTGGTTCGCTGGTGGGTCGCGAGCTCGATGGGAGCGCTTCCTCACCGCTACGCGCT 3370
 Qy 1029 -----GluMetProLeuGluProAspL 1036
 Db 3371 ACTAGCGCGACTGGTACGCCAGCGCGACTGGAGATCCAACTGCGCGCGCGAA----- 3423
 Qy 1036 euSerArgTyrLeuLysLysAspLysThrPheLysGlnLeuMetAlaSerHisGlyT 1056
 Db 3424 -----CACTGGCAGCGCATCTCTCGAGCCCTAC----- 3450

Qy 1056 hrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheA 1076
 Db 3451 -----GCCAAGCAGGCGAGCAGACCCCGCTGCTGCAGCGTGCCTGATCTCTGG 3499
 Qy 1076 laLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleA 1096
 Db 3500 CTTTC----- 3504
 Qy 1096 snAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValA 1116
 Db 3505 -----GCCCGCAGCATGCAATTGCGCGTGAACACCCCTGCTCGCGCGCTGCTCA 3553
 Qy 1116 spGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgAsgGluL 1136
 Db 3554 ACAT-----CTGGCCAAACCGCGCGGAGGAGGCGCGCGCGGAGC 3595
 Qy 1136 euLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyA 1156
 Db 3596 CGCTG-----GAGCCGACGACGCGCTGGTGC 3622
 Qy 1156 rgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleA 1176
 Db 3623 TGGCGCAGCCCGACTCGCGCTGCGCGTGGCGCGCGGAGTGTGCGGTGGACCTTG 3682
 Qy 1176 spLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAsp----- 1191
 Db 3683 CACGGCAACTGCGGGTGGCGGTGCGCGTGGCGCGCGCGCGCGAGCGCGAGCGGACGC 3742
 Qy 1192 -----ThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrThrP 1209
 Db 3743 AGCGCTGCTGAGCGCGCTTGGCGTTTCCGATGCTTGTGCGCTT----- 3792
 Qy 1209 hePheLysGluLeuLysSerAspLysSerArgSerSerAlaLeuLeu 1223
 Db 3793 -----CGCTCGCGCTGGATGGCCAGCAGCGCGCGCTCTGCTG 3831
 RESULT 14
 AAV05287
 ID AAV05287 standard; DNA; 49377 BP.
 XX AC AAV05287;
 XX DT 17-OCT-2003 (revised)
 DT 21-MAY-1998 (first entry)
 XX DE The soraphen biosynthesis gene cluster from Sorangium cellulosum.
 KW Polyketide synthase; PKS; biosynthesis; soraphen; SorR; SorA; SorB; SorM;
 biosynthetic module; beta-ketoacyl synthase; acyltransferase;
 ketoreductase; beta-ketone processing domain; cytostatic agent;
 antimicrobial agent; phytopathogenic fungi; transgenic plant;
 biological control; ss.
 XX OS Polyangium cellulosum.
 XX PH 383. .760
 FT CDS /tag= a
 FT /product= "SorR"
 FT /note= "gene product highly homologous to the reductase
 domains of type I PKSs such as eryA from
 Saccharopolyspora erythraea"
 FT CDS 927. .19874
 FT /tag= b
 FT /product= "SorA"
 FT /note= "gene product is highly homologous to type I PKSs
 that are known to be involved in the synthesis of
 polyketide compounds"
 FT misc_feature 942. .7115
 FT /tag= c
 FT /note= "module 1 of SorA"
 FT misc_feature 7203. .12884

FT /tag= d
 FT /note= "module 2 of SorA"
 FT 13455. 19616
 FT /tag= e
 FT /note= "module 3 of SorA"
 FT 19870. 24556
 FT /tag= h
 FT /note= "module 1 of SorB"
 FT 19871. 46318
 FT /tag= f
 FT 19871. 46318
 FT /tag= g
 FT /product= "SorB"
 FT /note= "gene product is highly homologous to type I PKs"
 FT genes
 FT 24638. 30820
 FT /tag= i
 FT /note= "module 2 of SorB"
 FT 30881. 35446
 FT /tag= j
 FT /note= "module 3 of SorB"
 FT 35528. 40114
 FT /tag= k
 FT /note= "module 4 of SorB"
 FT 40190. 46318
 FT /tag= l
 FT /note= "module 5 of SorB"
 FT 46851. 47891
 FT /tag= m
 FT /product= "SorM"
 FT /note= "gene product is homologous to the
 FT methyltransferase from Streptomyces hygroscopicus that is
 FT involved in the synthesis of the polyketide rapamycin"
 FT XX
 FT US5716849-A.
 FT PD
 FT 10-FEB-1998.
 FT XX
 FT 14-DEC-1996; 96US-00764233.
 FT XX
 FT 24-AUG-1993; 93WO-US007954.
 FT PR
 FT 08-JUN-1994; 94US-00258261.
 FT PR
 FT 09-OCT-1996; 96US-00729214.
 FT XX
 FT (NOVS) NOVARTIS FINANCE CORP.
 FT PA
 FT Ligon JW, Hill DS, Neff S, Beck JU, Ryals JA, Schupp T;
 FT WPI; 1998-158369/14.
 FT XX
 FT DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
 FT PT biosynthesis of soraphen useful as antimicrobial agent against
 FT PT phytopathogenic fungi.
 FT XX
 FT Claim 2; Col 47-90; 64pp; English.
 FT PS
 FT The present sequence contains a cluster of genes that encode polyketide
 FT CC synthases (PKSs) that are involved in the synthesis of soraphens in
 FT CC Sorangium cellulosum. The proteins encoded by the present sequence are
 FT CC SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
 FT CC which contain a beta-ketoacyl synthase, an acyltransferase, a
 FT CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
 FT CC processing domains. S. cellulosum soraphens are useful as a cyostatic
 FT CC and antimicrobial agent active against phytopathogenic fungi. Soraphen-
 FT CC producing transgenic plants or biological control agents can also be
 FT CC produced, which may reduce crop losses and nutritional deprivation for
 FT CC local populations in many parts of the world. (Updated on 17-OCT-2003 to
 FT CC standardise OS field)
 FT XX
 FT SQ Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 U; 0 Other;
 FT Alignment Scores:
 FT Pred. No.: 1.13 Length: 49377

Score: 192.00 Matches: 225
 Percent Similarity: 31.27% Conservative: 109
 Best Local Similarity: 21.07% Mismatches: 339
 Query Match: 2.60% Indels: 395
 DB: 2 Gaps: 51
 US-09-913-878A-2 (1-1402) x AAV05287 (1-49377)
 QY 5 ThrProArgLysArgAsnSerPro-----ValGlu 14
 DB 10696 TCGCTCGGTGCAAGCGCTCGCCAGCGCATCACGTCGCCGAGCAGCTCCGACCCCGAG 10755
 QY 15 GluIleAlaAsnArgLeuAsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThr 34
 DB 10756 GAGCTTCCACACCATGCTCTTCGCGTGGAGCTGGAGCGAGCTGCCGAGCC----- 10809
 QY 35 ThrLeuThrProHisArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHis 54
 DB 10810 -----CCACCTCACGCTCTGGAGCCCAAGGCGCTCTTCGATCGCGGCGCTCG 10863
 QY 55 AspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArgLysAspSerLeuAsnGln 74
 DB 10864 ACCTCGCCCGCGAGGTGCTCTCGCGCGTGGAGCTCGCTGCGCTCCAGAGCGCG 10923
 QY 75 AlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSerSerAsnTrpValProLysAla 94
 DB 10924 TCGACCAAGCGCTCGCTCCAGGCTCGTGTGCTCCCTTCATGGCTA----- 10974
 QY 95 HisAlaAspProAspThrLeu----- 101
 DB 10975 -----GAACCGCGAGCAGCTCATCCAGAGCGCCACTCCATCCGCGCGCGCTCG 11028
 QY 102 -----ProTrpSerLysGluProProArgAlaAlaThrAlaGly 114
 DB 11029 CCTGTGTCGAAGCTGGTGGCGCGAGCAACGCTCGCTCTCGCGCTCTCGCTCTCTCA 11089
 QY 115 GlnGlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProPro 134
 DB 11089 CCGGAC-----GGCCCATCGCTGCCGCGCG 11115
 QY 135 AsnAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArg 154
 DB 11116 ATGAAGACG---TCAAGGACCTCGCTACGCGCCCTCTCTGGGGGTTCGACGCTCCGCGC 11172
 QY 155 ProThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspPro 174
 DB 11173 AAAGCGACACCCAGACTCCGCTCTTCTCG---TCGACCTGGACCTCAGTAGGCGCT 11229
 QY 175 Pro-----LysArgSerLeuThrArg---SerAlaThr 184
 DB 11230 CCGACACACCTGCTCGCGCGCTCGAAACAGAGAGAGGTCACTCGCTCTCCGCAACG 11289
 QY 185 GlyProPro-----IleHis----- 189
 DB 11290 GAAACCTTCATCCGAGATTGGCGATGCAAGCTCGAAGATGAGCTCATCGCCCGCG 11349
 QY 190 -----GlyAlaAlaIleProLeuLysPhe---ProAspProValAsnThrGly 204
 DB 11350 ACGCGTCCAACTGGCGCTCCATA-----TTCGACCAAGGCAACTTCGACGCGC 11400
 QY 205 SerLys-----ArgProSerLeuGluSerGluAsnLeuAsn 216
 DB 11401 TCACCTCTCGACGCCCTCTAGCGCGTGGCGCCCTCGACACGCGCAAGTCCGCGTGC 11460
 QY 217 GlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaPro 236
 DB 11461 CCGTGCACG----- 11469
 QY 237 ProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArg 256
 DB 11470 CCGGACGCTCAATTTCCGCGATGCTCCGACACACCTTGGTGTGTATCCGGCGCAGCGG 11529
 QY 257 AspProThrAlaThrGlyHisArgAlaAspGlnValAspSerPheAspThrSerGln 276

Db 11530 GACCGCTCGGCG---GCGAAGCGCGAGCATGCTTACTGAAGTCGGTCCAGGTGTTTCGC 11586
Qy 277 GlyThrSer-----TyrGlySerSerValPheSerAlaCysArgHisAsn 291
Db 11587 GGTACACCGTAGCGACCGCGGTGATGCGGATCTTCGGCGCATGCGGTCCCGACGCCA 11646
Qy 292 GlnSerThrGlnSerSerPheGluAlaProProSerGlnProArgGluLysArgPro 311
Db 11647 TCGCCGACCGCGCATGATCTGCCCATCCCGACCGCTGCTTCCTTCGCCCAAGCCGCCA 11706
Qy 312 ValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSer-----LysGly 329
Db 11707 GCG-----TCCCATCATCTCTCATCCG 11730
Qy 330 ArgThrThrLysSerHisIleAspAsn-----GlnProLeuSerSerSerSer 345
Db 11731 CCTACTACGAGTCTGCTGATCTCGGGCATCTGAACCCCAATCAACGTCCTCATCCATG 11790
Qy 346 -----GlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
Db 11791 CGGCCCGCGCGCTCGGAGCGCGCGTTCAGCTCGCACCGCCACCTCGCGCGCAGG 11850
Qy 361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeu-----AlaArg 378
Db 11851 TCT-----TTGCCACCGCGCGCGAGGAGGAGTGGAGCGCTCTCCGCGCGC 11895
Qy 379 SerGluGluSerAlaArgSerGlnValHisAlaIleProValAlaAlaArgLe 398
Db 11896 TCGGCTTCGACGACGCGCACCTCG---CGTCTCAGTGCACCTGAGTTCGAGCAGCACT 11952
Qy 398 uArgAsnIleTyrProLysPheProLysTyrLeuHisGlu----- 411
Db 11953 T-----CCTGCGCTCCACGATGCGCGCGCGGTGATGCTCTCGACTGCT 12000
Qy 412 -----AlaProLeuAlaValAlaIleTyrGluValThrAr 422
Db 12001 TGSCACGCGAGTTCGTCGACGCTTCGTCGCGCTCATCGGAGCGGTGGAGCTTCGTCG 12060
Qy 422 g-LeuPheMetHisCysLysValAspLeu---GluAspGluSerLeuGlyLysTyrA 441
Db 12061 AGATG-----GGCAAGACGAGCATCCGTCGAGCCGACGCGTCCGCGTTCGCTACC 12111
Qy 441 spProSerTyrPheThrAlaArgAspValThrAspile----- 453
Db 12112 CTGGTGTGTTTACGCGCTTCGACCTCATAGAGCGCGGATCGAATAGACAGA 12171
Qy 453 ----- 453
Db 12172 TGCTCGACAGCTGCTCAGCTCTTCGAGCGCGGTGCGTTCGTCGCGCGCCATCAT 12231
Qy 454 --TyrLysThrLeuTyrArgLeuAspAlaPheArg-----G 465
Db 12232 CTTGGGACATCCGTATGCCCCCAGGCTTTTCGCGGCTCGCTCAGGCGCGCATGTG 12291
Qy 465 lYlys-----ProPheProGluLysProProAsnAspValPheValThrA 480
Db 12292 GGAAGTTCGTCTCACCATACCCGTCATAGACCCCGAGGACCGCTCTCATCAGG 12351
Qy 480 laMetThrGly----- 483
Db 12352 GAGGACCGGCACGCTAGAGCCCTGTGCGACGCCATCTCTGCGAAGACACGCGGCCA 12411
Qy 483 ----- 483
Db 12412 AGCACCTGCTTCTCAGTCGAGGCGCGCGCACGCTCCGGCGCGGAGGCTCGCAA 12471
Qy 484 --AsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyrAsnPro- 502
Db 12472 CCGAGTCGAAAGCGTGGGGGCTCTGTACACTTCGCGGTGCGACGCGCGCGCCAC 12531
Qy 503 -----AspAsnSerProThrAla---ProLeuTyrLeuValLysL 515

Db 12532 GCGCCCTCCAAGCCCTCTTGGACTCCATCCCGAGCGCTCACCGGCTCACCGCGCTCGTCC 12591
Qy 515 euLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgP 535
Db 12592 ACSCCGCGCGCGCTTCGACGACGCGC-----CTGCTCGCGCGCATGACCCCGACGCGA 12645
Qy 535 hePheGluIle-LeuIlePro-----SerProThrSer 545
Db 12646 TCGACCGCGTCTTTCGCCCAAGCTCGATCTGCTTGGACATTTGCATGAGCTCACCCAAG 12705
Qy 546 ThrSerProSerValProProValValSerLys----- 556
Db 12706 ACAAGCCCTTCGCGCGCTTCGCTCTCTCTGCGCTGCTGCGCTCTGCTGCTGCTGCTG 12765
Qy 557 -----GlnProGlyAlaValGluGluValIleGlnTrpLeu 568
Db 12766 GTCAGTCGAACACTACGCGCGCGCAATGCTTCTCGATGCGCTCGCGCATCACCGCGCTG 12825
Qy 569 ThrMetGly-----GlnHis-SerLeuValGlyArg- 578
Db 12826 CCACGCGCTCCCGCGCTCTCGCTCGCATGGGCTATTGGGCGGAGCGCAGTCGATGA 12885
Qy 579 -----GlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLys----- 592
Db 12886 CCGAGCAGCTCACGCGCGCGATGTTTTCGATGAGGCGCGCGCGCTCCGCGCCCTCG 12945
Qy 593 -----ProLeuArgGluPheGlnLeuArgAlaGlu-----As 603
Db 12946 CCACAGCAGGCGCTCTCCCTCTTCGATGCGCTCTTCGCGCGCGAGCGCGCCCTCG 13005
Qy 603 pProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheAr 623
Db 13006 TCCCGCAGCTTCGACGTGAACGCGCT-----CG 13035
Qy 623 gProAspValPheLysThrArgSerValProAlaGluGluProValGluGlnArgTh 643
Db 13036 GCGCAATGCGACGAGGTGCCCCCTGTTCAGCGCTTCGCTCCGCGTTCGCGTTCGCGAC 13095
Qy 643 rGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnThrTrpGl 663
Db 13096 GCAA-----GGCGCGCAGCAATACCG 13116
Qy 663 nProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMe 683
Db 13117 CCCTGGCTCTTCACTCTCTCAGCGCTCTCTCCCTCCGCGCGAGAAAGCAGCGCT 13176
Qy 683 t-----ThrLeuGluProHisGlnIleArgHisHisLysThrAspLeu----- 698
Db 13177 TCCTCTCGATCTCGTCGACCGAAGCGCCACCGCTCTCGGCTCGCTCATTCGAAT 13236
Qy 699 -----SerProSerGlyThr-GlyGluValMetAsnAspGlyVal----- 711
Db 13237 CGCTCGATCCCATCGCCCTCTCAAAGAGCTCGCGCTCGATTCTCTATCGCTCTCGAGC 13296
Qy 712 ---GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
Db 13297 TCGAGTTCAGTTCG 13353
Qy 730 spValProSer---AlaValGlnGlyArgPhe----- 739
Db 13354 ACTATCCAAACCGCGTGCATCTCACGCTTTTTCACGACGCGAGTCTTTCGGGGAACCA 13413
Qy 740 -----GlySer----- 741
Db 13414 CCGACCGTCCGCGAGCGCGCTCACCCCGCGGGAAGCGAAGACCTATCGCCATCGTGT 13473
Qy 742 -----AlaLysGlyMetTrpValIleAspV 750
Db 13474 CGATGACCTCGCTTCCCTGGCGCGAGTGGCGACCGCGCGAGGATCTCTGGAAGCTCTTCG 13533
Qy 750 alAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluC 770
Db 13534 TCGAT-----GGGAAGATGCCATCTCCAGCTTTCGCCAGAAATCGCGGTGGAGTC 13584

```

QY 770 yAspPheValasp 774
Db 13585 TCGATCGCTCGAC 13598

RESULT 15
ID AAZ90112 standard; cDNA; 4148 bp.
XX AAZ90112;
XX AC
XX 19-MAY-2000 (first entry)
XX Human antizuai-2 (AZ-2) gene.
XX Antizuai-2; AZ-2; human; breast cancer; tumour suppressor; ss;
XX malignancy progression marker; malignancy reversion; chromosome 10q26.
XX Homo sapiens.
XX WO200000503-A1.
XX 06-JAN-2000.
XX 25-JUN-1999; 99WO-US014482.
XX 26-JUN-1998; 98US-0090747P.
XX (CHEN/) CHEN H.
XX (SIS/) BISSELL M.
XX Chen H, Bissell M;
XX WPI; 2000-170903/15.
XX P-PSDB; AA78795.
XX New nucleic acid encoding a tumor suppressor or marker, used for
XX diagnosis, monitoring progress or treatment, and gene therapy of breast
XX cancer.
XX Claim 2; Fig 5; 120bp; English.
XX This sequence represents the human antizuai-2 (AZ-2) gene sequence. AZ-2
XX is a variant of the antizuai-1 (AZ-1) gene (see AA290111). The AZ-1 gene
XX is located on chromosome 10q26, and encodes a protein that acts as a
XX tumour suppressor or marker of malignancy progression or reversion. The
XX AZ-1 protein and its variants are tumour suppressors, AZ-1 interacts with
XX E-cadherin and beta-catenin. Detecting low levels of AZ-1 nucleotide or
XX amino acid sequences are used to diagnose a breast cell malignancy, also
XX for monitoring disease progression, particularly assessment of
XX therapeutic efficacy. The nucleotide sequence is used in vivo or ex
XX vivo gene therapy, and AZ-1 polypeptides are used for treating or
XX preventing breast cancer. AZ-1 polypeptides are also used to raise
XX specific antibodies, for diagnostic detection of AZ-1. Fragments of the
XX AZ-1 nucleotide sequence are useful as probes or primers for detecting
XX expression of the AZ-1 gene
XX
XX SQ Sequence 4148 BP; 1107 A; 1140 C; 1082 G; 819 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.0944 Length: 4148
Score: 188.50 Matches: 197
Percent Similarity: 33.95% Conservative: 154
Best Local Similarity: 19.05% Mismatches: 386
Query Match: 2.55% Indels: 298
Db: 3 Gaps: 45

US-09-913-878A-2 (1-1402) x AA290112 (1-4148)

QY 1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleAsnArgLeu 20
Db 676 ATGTTACCTTCGGTTCCT--AAGAGGATGCTCCA-----AGAGTCATGGATAAAGTC 726

```

| | | | | | | |
|----|------|---|------|------|-----------------|------|
| Db | 1639 | ---- | | ---- | CCCGCTGGGGCCAAA | 1655 |
| Qy | 365 | GLValaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGluGluSerAlaArg | 384 | | | |
| Db | 1657 | GCtGCCTCCCTCTGGACTCAGAGCTCAGAGGGGTTCTCCCGCGCTTCTGAGGT | 1716 | | | |
| Qy | 385 | SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys | 404 | | | |
| Db | 1717 | GGCAGAGTCAGAACTCACCCTCTGCGGAGGAAACCGTG | 1758 | | | |
| Qy | 405 | PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe | 424 | | | |
| Db | 1759 | -----CCTCTTACCAGCCCGGAGGAGGAGGAGTAACC | 1794 | | | |
| Qy | 425 | MetHisCysLysValAspLeuGluAspGlu | 439 | | | |
| Db | 1795 | CCATCGGATAGCGGGGCAAGAGGACTCCAGCCAAAGGGCTCTCCGTAAGGCTGGAG | 1854 | | | |
| Qy | 440 | TyrAsp-----ProSerTrpSerThrAlaArgAsp | 449 | | | |
| Db | 1855 | TTTGACTATTCTGAGGACAAGTAGTTGGGACAACGAGCAGGAGAACCCCTCTCTACC | 1914 | | | |
| Qy | 450 | -----ValThrAspIleTrpLys | 455 | | | |
| Db | 1915 | AAAAAGATAGCAAAAAGCCAGTTGCCAAATGCCCTGAGGAGGCCAAAGATGAAAAAG | 1974 | | | |
| Qy | 456 | ThrLeuTyrArgLeuAspAlaPheArgGlyIysProPheProGluLysPro | 473 | | | |
| Db | 1975 | ACCCCGAGAAACTTGACAACACTCTCGCTCACTCTCCAGATCCCTGTGTAACCC | 2031 | | | |
| Qy | 474 | AsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaValVal | 493 | | | |
| Db | 2032 | AATGACATCCCATGTCTAAAGGTACTTACACCTTTGAT | 2070 | | | |
| Qy | 494 | LeuSerAlaValLeuAspTyr-----AsnProAspAsnSerProThrAlaProLeuTyr | 511 | | | |
| Db | 2071 | -----ATTGACAAAGTGGGATGACCCCAATTTTAAACCTTTTTCACCTCA | 2118 | | | |
| Qy | 512 | LeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArgAspGheGly | 531 | | | |
| Db | 2119 | AAATGACAGGAGTCTCCAAACTGCCCAACAA | 2166 | | | |
| Qy | 532 | ProAspArgPheGluIleLeuPro-----SerProThrSerThrSer | 547 | | | |
| Db | 2167 | CCAGACACCTGTGATGAGTCGTTGACCCCTTTAAGACATCCTTAAGACCCCGAGTCA | 2226 | | | |
| Qy | 548 | ProSerValProProValValSerLysGlnProGlyAlaValGluGluVal-IleGlnTr | 567 | | | |
| Db | 2227 | CCITCTAAATCCCGAGCTCTTTGATGCCGCCAGTCTGTGAGGAGCAATGAGTG | 2286 | | | |
| Qy | 567 | pLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhe---PheAlaLys | 586 | | | |
| Db | 2287 | GACGGGATGGCTAAACAAGCCGCCCAAGAGAAGAACGCCCTCTAAGACGGTGAA | 2345 | | | |
| Qy | 586 | sAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspPro | 604 | | | |
| Db | 2346 | AAAGTCGCAAAACAGGCTCTCTCTCTGAT-----CCACCTTCCGAGACCCCAACC | 2399 | | | |
| Qy | 605 | -----LysProIleIleLysGluArgValHisPheAlaGluThr | 618 | | | |
| Db | 2400 | AGCTGCTACACCAAGAACACACAGTGAATCTCGCGTGTCTCCAGCCACATGAGGA | 2459 | | | |
| Qy | 618 | rGlyLe----- | 620 | | | |
| Db | 2460 | AAAGTCGGGTTCACCAACCAAGAGTAGGAGTGCATGACAGTGGACCTAGAGGCTGACAA | 2519 | | | |
| Qy | 621 | -----ThrPheArgProAspValPheLysThrAr | 630 | | | |
| Db | 2520 | ACAGGACTACCGCAGCGCTCGGACCTGTCCACCTTT-----GTAAACGAGACCAA | 2570 | | | |
| Qy | 630 | gSerValValProAlaGluProValGluGlnArgThrGluPheLysValSerGlnMe | 650 | | | |

Search completed: March 31, 2004, 07:25:55
Job time : 1026.39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 06:31:56 ; Search time 161.141 Seconds
(without alignments)
4928.317 Million cell updates/sec

Title: US-09-913-878A-2
Perfect score: 7397
Sequence: 1 MNPTPRKRNFSVEINRL.....YEVIGDDDFDGIQGTNGDY 1402

Scoring table:
BLOSUMP2
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USFTO_spool/US09913878/runat_29032004_082009_8395/app_query.fasta_1.2254
-DB=issued Patents NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANSHUMAN40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878 @CGN 1 1.82 @runat_29032004_082009_8395 -NCPUS=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 297.5 | 4.0 | 3731 | 3 | US-08-811-583-1 |
| 2 | 192 | 2.6 | 4377 | 1 | US-08-764-233A-1 |
| 3 | 182 | 2.5 | 4026 | 4 | US-09-252-991A-10214 |
| 4 | 178.5 | 2.4 | 28958 | 1 | US-08-258-261B-6 |
| 5 | 178.5 | 2.4 | 28958 | 1 | US-08-456-837-6 |
| 6 | 178.5 | 2.4 | 28958 | 1 | US-08-457-342-6 |
| 7 | 178.5 | 2.4 | 28958 | 1 | US-08-457-646A-6 |
| 8 | 178.5 | 2.4 | 28958 | 1 | US-08-458-076A-6 |
| 9 | 178.5 | 2.4 | 28958 | 1 | US-08-764-233A-4 |
| 10 | 178.5 | 2.4 | 28958 | 1 | US-08-457-335A-6 |
| 11 | 178.5 | 2.4 | 28958 | 1 | US-08-729-214-6 |
| 12 | 178.5 | 2.4 | 28958 | 3 | US-09-028-934-6 |

| | | | | | | |
|------|-------|-----|---------|---|---------------------|--------------------|
| 13 | 174 | 2.4 | 14770 | 4 | US-09-220-132-30 | Sequence 30, Appli |
| C 14 | 171 | 2.3 | 52297 | 3 | US-09-426-436-1 | Sequence 1, Appli |
| C 15 | 171 | 2.3 | 52297 | 3 | US-08-705-557-1 | Sequence 1, Appli |
| C 16 | 162 | 2.2 | 15872 | 3 | US-09-105-537-1 | Sequence 1, Appli |
| C 17 | 162 | 2.2 | 15872 | 4 | US-09-091-609-1 | Sequence 1, Appli |
| C 18 | 162 | 2.2 | 15872 | 4 | US-09-091-609-3 | Sequence 3, Appli |
| C 19 | 160.5 | 2.2 | 2943 | 1 | US-08-042-747A-7 | Sequence 7, Appli |
| C 20 | 160 | 2.2 | 8312 | 4 | US-09-620-312D-1048 | Sequence 1048, Ap |
| C 21 | 158 | 2.1 | 2419 | 4 | US-09-624-693A-12 | Sequence 12, Appli |
| C 22 | 158 | 2.1 | 77536 | 4 | US-09-410-551B-1 | Sequence 1, Appli |
| C 23 | 158 | 2.1 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appli |
| C 24 | 156.5 | 2.1 | 2472 | 4 | US-09-252-991A-886 | Sequence 886, App |
| C 25 | 156.5 | 2.1 | 2955 | 4 | US-09-252-991A-797 | Sequence 797, App |
| C 26 | 156 | 2.1 | 13987 | 2 | US-08-804-227C-13 | Sequence 13, Appli |
| C 27 | 156 | 2.1 | 44377 | 2 | US-08-804-227C-7 | Sequence 7, Appli |
| C 28 | 156 | 2.1 | 44377 | 2 | US-08-804-198-1 | Sequence 1, Appli |
| C 29 | 156 | 2.1 | 4403765 | 3 | US-09-103-840A-2 | Sequence 2, Appli |
| C 30 | 155.5 | 2.1 | 4403765 | 3 | US-09-103-840A-2 | Sequence 2, Appli |
| C 31 | 155.5 | 2.1 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appli |
| C 32 | 155 | 2.1 | 11907 | 3 | US-08-061-376-4 | Sequence 4, Appli |
| C 33 | 154.5 | 2.1 | 77536 | 4 | US-09-410-551B-1 | Sequence 1, Appli |
| C 34 | 154 | 2.1 | 30001 | 1 | US-08-125-468-1 | Sequence 1, Appli |
| C 35 | 154 | 2.1 | 30001 | 2 | US-08-474-933-1 | Sequence 1, Appli |
| C 36 | 154 | 2.1 | 33529 | 3 | US-09-144-085-3 | Sequence 3, Appli |
| C 37 | 153.5 | 2.1 | 11283 | 2 | US-08-603-753D-3 | Sequence 3, Appli |
| C 38 | 153.5 | 2.1 | 11283 | 3 | US-09-099-753-3 | Sequence 3, Appli |
| C 39 | 153.5 | 2.1 | 11283 | 3 | US-08-986-108-3 | Sequence 3, Appli |
| C 40 | 153.5 | 2.1 | 13613 | 3 | US-09-105-537-3 | Sequence 3, Appli |
| C 41 | 152.5 | 2.1 | 2941 | 3 | US-09-425-383-1 | Sequence 1, Appli |
| C 42 | 152.5 | 2.1 | 36519 | 3 | US-08-923-137-2 | Sequence 2, Appli |
| C 43 | 151.5 | 2.0 | 11385 | 2 | US-08-639-501-1 | Sequence 1, Appli |
| C 44 | 151.5 | 2.0 | 11385 | 3 | US-09-044-946-1 | Sequence 1, Appli |
| C 45 | 151.5 | 2.0 | 11385 | 3 | US-09-044-908-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1

US-08-811-583-1
; Sequence 1, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenegeger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (R&RP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSES: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MFG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090

; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3731 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Tomato
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 194..3535
 ; US-08-811-583-1

Alignment Scores:

Pred. No.: 3,73e-16 Length: 3731
 Score: 297.50 Matches: 189
 Percent Similarity: 35.79% Conservativeness: 137
 Best Local Similarity: 20.75% Mismatches: 339
 Query Match: 4.02% Indels: 247
 DB: 3 Gaps: 35

US-09-913-878a-2 (1-1402) x US-08-811-583-1 (1-3731)

QY 562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
 DB 1349 GAAGACATAGATACTTCTTCGTGT-----TCTTTTGTGATGAGAGTGGAG 1399
 QY 582 AlaPheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
 DB 1400 AAATCTGATTTACAGACTTA----- 1420
 QY 602 GluAspProLysProIleIleLysGluArgValHisPheAlaGluThrGlyIleThr 621
 DB 1421 ----TTACCAAAA-----GCCAAGTCTGGAAGTGGT 1447
 QY 622 PheArgProAspValPheLys-----ThrArgSerValValProAla 635
 DB 1448 CTCAGGACAAACATCTATGAGAGGATCTTATCACTCTCGGAAAGGCTTGTAAATGGT 1507
 QY 636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
 DB 1508 GATAAAAAATTGAA---TTTCTTCATTTTCATCGAGCCAGTTG--- 1549
 QY 656 GlnLeuAspAsnAsnThrTrp-----GlnProHisLeuLysLeu----- 668
 DB 1550 ---CGGGATAATTCAGTGTGGATGTTGCATCAAGACCTGGCTTACTGCACAAATGATATA 1606
 QY 669 -----PheSerArgIle-----GlnLeu 674
 DB 1607 AGAGCTTGGATGGGTGATTTTTCGACAGATCAAGAATGTCCGCAAAATATGCTGCCAGACTT 1666
 QY 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLys 694
 DB 1667 GGTCAATCTTTGGTTCCTCCAGAGAGACTTGG-----AGTGTCTTAGGCATGAG 1717
 QY 695 ThrAspLeuLeu-----SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
 DB 1718 ATTGAAGTATTCCCGATGTAAGGTTTCATGGAACCCAGCTATGCTCTTCTGATGGAATT 1777
 QY 712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
 DB 1778 GGTAAATCTCGTGACTTGTCTATAGAGTTGGCTCAAAATGTCGCTTCAATATACC 1837
 QY 732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
 DB 1838 CCATCTGCTTCCAGATTCTGTTATGGTGGATATAAAGGTGTGTGGGTGTTGATCCGGAT 1897
 QY 752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
 DB 1898 TCATCAATGAAG-----TTGCTTTTGAGAAAGAGCATCTCGAAATATGAATCAGAC 1948
 QY 772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791

DB 1949 AACATA-----AAAGTTAGATGTCCTTGA---TGAGAGCAAAATATCAGCCT 1990
 QY 792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
 DB 1991 TGTATCTTAATCGTCACTGATTACGCTCTGTCT----- 2026
 QY 812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
 DB 2027 ---ACACTTGGAGTGAAGATGAAGTTCTCGAACAGAACAGCAAGAGCTGTAGATCAG 2083
 QY 832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
 DB 2084 CTTGATGCTATCTTGCATGATCTTTTGAAGGCACAGGAGCTTTGGAATTGATGCT--- 2140
 QY 852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
 DB 2141 -----CCTGGAGAGAAC 2152
 QY 872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
 DB 2153 ACTAATATTCTCAAGCAATGCTAAACTGTGTATTAAAGCCTGATGCTGAGCCCTTCTT 2212
 QY 892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
 DB 2213 TCATGATGTTGCAAACTTCGCGCATCCAAAGTTGCTCGATTTCGGACTAGATCAAGA 2272
 QY 912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
 DB 2273 ATATTTATTTCCAAATCGAAGAACAAATGATGGGATGTTGGATGAATCCAGAACCTTGGAA 2332
 QY 932 GluAsnGluValHisValGlyPheSer-----SerLysPheArgAspGluGlu 947
 DB 2333 TATGGTCAGGTGTTGTTTCAGTTTACTGTTGCTGGACATGGAGAGTTTCTGACGATTGA 2392
 QY 948 GluSerPhe-----ThrLeuLeuSerAspCysAspValLeu 959
 DB 2393 CATCCATTAAATACAGCAGATCCACCAACAGTAATTTCTTCCTGAAGGAATATGGGT 2452
 QY 960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
 DB 2453 GTTGCAAAAAATCCATGCTGTGATCTCTGATGATTTCTGTTTAAAGCTGTAAATGTT 2512
 QY 980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
 DB 2513 CGAGCGCTGCACCATGTTAGATTGTTGTTATTCCTCCAGAAAGAGAAAGACCTCAT 2572
 QY 1000 AlaLysLysLeuSerGlyLysAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
 DB 2573 CCGAATGAATGTTCTGGAGTGAATTTGGATGGGATATCTACTTTGTTGCTGGATCAA 2632
 QY 1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
 DB 2633 GACATGATCCCGCAAGGCAAGTCCAGCGATGGAATATCTCCAGCACCACCATACAG 2692
 QY 1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
 DB 2693 TTGGACCATGAT----- 2704
 QY 1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
 DB 2705 -----GTCACAATTGAGGAAGTTGAAGAGTACTTCCACCACTATATTGTGAAT 2752
 QY 1080 AsnPheLeuGlyMetCysThrAsn-----TyrLysGluArgLeuCysTyrIle 1095
 DB 2753 GACAGTTTGGGAATCATAGCAAAATGCCATGTCGTATTTCGACACAGA-----GAA 2803
 QY 1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
 DB 2804 CCTGATATGCCCATGATGATCCATCAAAAAAATTGCTGAGCTCTTTTCAATTGCAGTG 2863
 QY 1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgAspGlu 1135

Db 2864 GACTTCCAAAGACTGGTGTCCCGCTGAATACCATCT-----CAGTTGCGC----- 2911
Qy 1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrrLysserAsp-SerTrpLeuG1 1155
Db 2912 -----CCTAAGAATACCAGACTTCATGATAAG 2941
Qy 1155 yArgGlyGluProThrHisIlelleAspTyrrLeuLysPheSerIleAlaArgProAlaI1 1175
Db 2942 CCGG-----ACAAAGACCAGCTAT 2959
Qy 1175 easPlysglu-----LeuGluAlaPheHisAsnAlaMetIlysalalalysAspThrG1 1193
Db 2960 ATCTCAGAAAGAGTTATTGGAAGCTTTTC-AGGAAAGTGAAG----- 3001
Qy 1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrrThrPhePheLysGluI1 1213
Db 3002 -GACAAAGCA-----CCTCAGGCTAGCTCTATCCGACCTTCACAGAGATG 3048
Qy 1213 eSerAspLysserArgSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyG1 1233
Db 3049 TGCAAGGAGATCATATGATGCTGATGGAAGTTGATGATTTGAAGATTACATTGACGA 3108
Qy 1233 uVal-----GluLysGluTyrrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
Db 3109 AGCTTTTGAACAAACTGATAT-----CACAAAGCTGGGTATTT 3153
Qy 1250 rLysAspProTyrrProValArgValAsnGlnValTyrrGluLysTrpCysAlaIleThrPr 1270
Db 3154 AATGACTACTATGTCATATAAAGACAGAG-----GCTGAATACTATTAG 3195
Qy 1270 oGluAlaMetAspLysserGlyAlaAsnTyrrAsp-----Se 1282
Db 3196 TGGTGGCATTTAAGGAGCATCAAAACTTTTGCCGCAAGAAAGATCGTAGGCCANTAG 3255
Qy 1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
Db 3256 TGTTCTGTGAGGCTTTGAGG-----AAGGAGGCAAGAGCCTGCTT 3297
Qy 1302 -----AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
Db 3298 CAAGAGGCGTAAATGATATAGATGACATGTTACCAAGGCTTCGGCTGGTACCAGTTAC 3357
Qy 1313 rTyrrHisLysser-----ProLy 1319
Db 3358 ATATCATCTACATATGCGGTGCTACATCAGGGTTGAAAGAGCTCATTTTCATTAG 3417
Qy 1319 ePheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrrIleLysAlaGlnMetThrSerAr 1339
Db 3418 CTTTCCCTGGTGTGTTTATGACCACTAATCCAGATTAAAGAAAGCAACAGCACTAACAG 3477
Qy 1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrrAlaGlyLeuMetProAs 1359
Db 3478 GCCAGTTCTCACTGTCTCTCTCAGGGCTCACTGAGTCACAGATTAGTG----- 3529
Qy 1359 pLysLysPheThrLysGlnTyrrValAlaArgLeuGluAspGlySerGluTyrrProAs 1379
Db 3530 -----TTGAAATGAGATTCCAGTCGAGCGTTTAAAG 3558
Qy 1379 pProGluValTyrrGluValLeuLysAspAsp 1389
Db 3559 CTGATATATATATATGTAATGATAGGTGTGAT 3589
RESULT 2
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE: Sorangium cellulosum
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pJL3, and pVKM15
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSs such as eryA from
; OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs the
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of Sora"
; OTHER INFORMATION: /note= "Module 1 of Sora"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product= "Module 2 of Sora"
; OTHER INFORMATION: /note= "Module 2 of Sora"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product= "Module 3 of Sora"
; OTHER INFORMATION: /note= "Module 3 of Sora"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19871..46318
; OTHER INFORMATION: /product= "Sorb"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS gene

Db 12061 AGATG-----GGCAAGACGGACATCCGTGAGCCGCGCGTGGCGTCCGTACC 12111
 Qy 441 spProSerTrpSerThrAlaArgAspValThrAspIle----- 453
 Db 12112 CTGGTGTGGTTTACCGCGCTTCGACCTCAATAGAGCGCGGAGCGGATCGAATAGAGAGA 12171
 Qy 453 ----- 453
 Db 12172 TGCTCGCAGAGTGTCTCAGCTCTTCGAGCGGGTGGCTTCGTCCGCGCCCATCAT 12231
 Qy 454 ---TrrpLysThrLeuTrpArgLeuAspAlaPheArg-----G 465
 Db 12232 CTGGGACATCCGTATGCCCCCGAGCGCTTCGCGCGCTCGCTCAGCGCGCGCATGTG 12291
 Qy 465 lYlys-----ProPheProGluLysProAsnAspValPheValThrA 480
 Db 12292 GGAAGTGTCTCTCACCATACCCCTCCCATAGACCCCGAGGACCGCTCTCATACGG 12351
 Qy 480 laMetThrGly----- 483
 Db 12352 GAGGACCGGCACGTAGAGCCCTGGTGTGCAAGCCATCTCTGTCGAAGACACGCGGCCA 12411
 Qy 483 ----- 483
 Db 12412 AGCACTGCTTCTACGTGAGGAGGCGCGCAGCTCCGCGCGCGAGCTCGCGAA 12471
 Qy 484 ---AsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyrAsnPro- 502
 Db 12472 CGAGCTCGAAGCGTGGGGCTCTGTACACTTCGCGCGTGGGAGCGCGCCGACCCAC 12531
 Qy 503 -----AspAsnSerProThrAla---ProLeuTyrLeuValLysL 515
 Db 12532 GCGCCCTCAGCCCTCTTGACTCCATCCGAGGCTCACCGCTACCGCGCTCGTCC 12591
 Qy 515 euLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgp 535
 Db 12592 AGCGCGGGGCGCCCTCAGCAGCGC-----CTGCTCGCGCCATGAGCGCCGAGCGCA 12645
 Qy 535 hePheGluLeu-LeuIlePro-----SerProThrSer 545
 Db 12646 TCAGCGGCTCTTCCCGCCAGCTGATGCTGATGCTGTCATCTGTCATGAGCTACCCCAAG 12705
 Qy 546 ThrSerProSerValProValValSerLys----- 556
 Db 12706 ACAAGCCCTCGCCGCTCTGCTCTCTCTGCTGCTGGCTGCTGCTGGTAGTCCAG 12765
 Qy 557 -----GlnProGlyAlaValGluValIleGlnTrpLeu 568
 Db 12766 GTCACTGAACTACGCGCGGCCAATGCTTCTCGATGCGCTCGCGCATCAGCGCGTG 12825
 Qy 569 ThrMetGly-----GlnHis-SerLeuValGlyArg----- 578
 Db 12826 CCCACGGCTCCCGGCTCTCGCTCGCATGGGCTATGGCGCGAGCGCAGTCAATGA 12885
 Qy 579 -----GlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLys----- 592
 Db 12886 CCGAGCACTCAGCGCGCCGATGTTCTCGCATGAGCGCGCGGCTCGCGCCCTCG 12945
 Qy 593 -----ProLeuArgGluPheGlnLeuArgAlaGlu-----As 603
 Db 12946 CCACAGACGAGCGCTCTCTCTCTGATGGGCTCTCTCGCGCGCGAGCGCCCTGG 13005
 Qy 603 pProLysProLeuLysGluArgValHisPheAlaGluThrGlyLeuThrPheAr 623
 Db 13006 TCCCGCACGCTTCCAGCTGAACGGCT-----CG 13035
 Qy 623 qProAspValPheLysThrArgSerValProAlaGluProValGluGlnArgTh 643
 Db 13036 GCGCAATGCCAGAGGTGCGCGCTGTTCCAGCGTCTCGTCCGCGCTCGCGTCCGAC 13095
 Qy 643 rGluPheLysValSerGlnMetLeuAspTrpLeuGlnLeuAspAsnAnthrTrpG1 663
 Db 13096 GCAG-----GGCGCCGAGCAATACCG 13116

Qy 663 nProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMe 683
 Db 13117 CCTTGGCTTCTCACTCTCTCAGCGCTCTCTCTCCCTCCCGCGAGAAAGCGAGCGCT 13176
 Qy 683 t-----ThrLeuGluProHisGlnIleArgHisLysThrAspLeuLeu----- 698
 Db 13177 TCCTCTCGATCTCTGCGCGAGCGCGCGCGCTCTCTCGCGCTCTCTCTCATTCGAAT 13236
 Qy 699 -----SerProSerGlyThr-GlyGluValMetAsnAspGlyVal----- 711
 Db 13237 CGTCTCGATCCCATCCCTCTCCAGAGCTCGGCTCGATCTCTTATTCGCTCTCGAGC 13296
 Qy 712 -----GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
 Db 13297 TCCGAGTCTGACTCGCGCGCGCGCGCGCTCTGAGCTCCAACTACTCTC---CTCTTCG 13353
 Qy 730 spValProSer---AlaValGlnGlyArgPhe----- 739
 Db 13354 ACTATCAACCGCGCTGCACTCTCAGCTTTTTCACGAGCGAGTCTTCTCGGGGAAACCA 13413
 Qy 740 -----GlySer----- 741
 Db 13414 CGAGCGTCCGCGAGCGCGCTCACCCCGCGGAGCGAAGACCCATATCGCATCGTGT 13473
 Qy 742 -----AlalysGlyMetTrpValIleAspV 750
 Db 13474 CGATGAGTCCCGCTTCTCTCGCGAGTCTCGCGAGCGCGCGAGTCTCTGGAAGCTCTTCG 13533
 Qy 750 alAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluC 770
 Db 13534 TCGAT-----GGGAAAGATGCCATCTCCAGCTTTCGCCAGAAATCGCGTGGAGTC 13584
 Qy 770 ysAspPheValAsp 774
 Db 13585 TCGATCGCTCGAC 13598

RESULT 3

US-09-252-991A-10214
 ; Sequence 10214, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: ASRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10214
 ; LENGTH: 4026
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10214

Alignment Scores:
 Pred. No.: 9,92e-06 Length: 4026
 Score: 182.00 Matches: 273
 Percent Similarity: 33.31% Conservative: 192
 Best Local Similarity: 19.56% Mismatches: 473
 Query Match: 2.46% Indels: 465
 DB: 4 Gaps: 65

US-09-913-878A-2 (1-1402) x US-09-252-991A-10214 (1-4026)

Qy 86 AlaSerSerAsnTrp-ValProLysAlaHisAlaAspProAspThrLeuProTrpSerly 105
 Db 401 TCGAAGCGCGTGGAGTTCTCTCAAGCGCGAGAAACCTGC-ACCGGCTCTCGCCGAG 459

QY 105 sGluProProArgAlaAlaThrAlaGlyGlnTrpAlaLeuGlnThrValLeuLeuG1 125
: : : : :
Db 460 GAGAGTTCAAGGGCAAGGCTGTGCAACACCTGGCTATCTCT 505
QY 125 uValLeuAsnArgPheMetProProAsnAsnThrProGlyArgThrPheGlyArgTh 145
: : : : :
Db 506 : : : : : 525
QY 145 rLeuSerGlyProSerGlyLeuSerArgProProThrSerThrAsnThrLysArgLysAspG1 165
: : : : :
Db 526 TCGCCCGGGTGATGACGCGCGCTTCTCTACGAGTCGCGCAACAGGTCACCGAGCGG 585
QY 165 uProAlaAsnValThrPheAlaAspPro---ProLysArg-SerLeuThrArgSerAlaT 184
: : : : :
Db 586 GTGCCCGAGCTAAGATGGGCAACGCCATGACCGCGCGCGCTACGACGCGCAACCG 645
QY 184 hr-----GlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspP 200
Db 646 CAATACGCGCGGATTCGCGCGCTGCGGTGCTACGACGAGTTCGCTA-----TCCG 696
QY 200 roValAsnThrGlySerLysArgProSerLeuGluSerGluAsn-----LeuAsnG 217
: : : : :
Db 697 CTGTGGGACGCCAAGCCATGACCGCGCGGAGGAGTCAAGCTGGCGGTCTCTCCAGC 756
QY 217 InCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaProp 237
: : : : :
Db 757 GAGTTCATCAACGCTGCGCGGCAATGCTACATCCGCTGGCAAGCTGAAGCCGGCG 816
QY 237 roValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgA 257
: : : : :
Db 817 CTGTACCTGTGTCGAGCGCTGTGCGCAAGTACCG---CGCCACACCGTGTGCTGTC 873
QY 257 spProThrAlaThrGlyHisArg---ArgAlaAspGlnValAspSerPheAspThrSerG 276
: : : : :
Db 874 TCCACACGCTGGCGGTGAGCAAGTCCCGCGGACGAGCTGGTCTGACCGCACGC 933
QY 276 InGlyThrSerTyGlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrg 296
: : : : :
Db 934 AA-----GCACGAAGCGCACCCCGCGTGGCGCGACCC 963
QY 296 InSerSerPheGluAlaProProSer-----GlnProArgGluLysArgP 311
: : : : :
Db 964 AAGGTGCTGTGGACCGCGCTGGCGTGATGACGAGCGGCATACCGCGCGCGCGC 1023
QY 311 roValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgT 331
: : : : :
Db 1024 CTGTGCGGCTCAAGCAGCGCAGTCCGAGCGCTCTACGT-----GATCGGCG-AGGA 1076
QY 331 hrThrLysSerHisLeAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerP 351
: : : : :
Db 1077 CCGCAGG-----GCGGGTCTTCTCTCCGAGAACTTCTA 1112
QY 351 heSerThrTyTrpGluSerPheProSerSerGlyGlyGluAlaIleProGluPro- 370
: : : : :
Db 1113 TTACGACACGAGATCTACGACACCAAGATCT-----ACGCTTTCACCGA 1157
QY 371 --SerArgSerAsnGlyLeuAlaArg-----SerGluGluSerAlaA 384
: : : : :
Db 1158 CGCCCGCTGTACCGCGCGGAGTCTGCTGTTGAGATGCTGCGCGCGAGTTCAA 1217
QY 384 rGSerGlnValGlnValHisAlaPro-----ValV 394
: : : : :
Db 1218 GACGCGCGGCAATCCAGCGCGCGCGCAGCGCGCGTGGCGCTGAGCGGTGTCGACGC 1277
QY 394 AlAlaAlaLeuArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProL 414
: : : : :
Db 1278 CAGCGCGAGGATTGCAAGCTGACCTCC---GCTTCAGCGGAGAGCGCGCCAA 1334
QY 414 euAlaValAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspG 434
: : : : :
Db 1335 TGGCGGCTTCC-----AACTGCGGAG-AAACCGCGTCCCG 1369
QY 434 luSerLeuGlyLeuLysTyArgP-----ProSerTrpSerThrAlaArgAspValT 451

Db 1370 GCGGCTACAGTGGCTTCGACTACCGCGGCGACACCTACAGCAGCGCTTCCGCGTGG 1429
: : : : :
QY 451 hrAspIleTrpLysThrLeuTyArgLeuAspAlaPheArgGlyLysPro----- 467
: : : : :
Db 1430 CCGAATACATCAAGCGCACCTTCGAAGTCGCTCGACCTGCGCAAGCGGACTTCAAGA 1489
: : : : :
QY 468 -----PheProGluLysProPro----- 473
: : : : :
Db 1490 CCGCGAGCGCGTGAAGGCGGAGATCGTCTCTCTATCCGACGCGCAAGCGTGGCCA 1549
: : : : :
QY 474 -----AsnAspValP 477
: : : : :
Db 1550 ACGCGCGCTGCAACTGAGCTCGCGGCCAGCTGTGCGATGTGTGGACACAGAACTG- 1608
: : : : :
QY 477 heValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSer----- 495
: : : : :
Db 1609 -----CAGTATCTCGCGCAGTTCCCGGTGCGAATGAGCAGCAGCGAACTGACCA 1660
: : : : :
QY 496 -----AlaValLeuAspTyArgAsnProAspAsnSerProThr----- 507
: : : : :
Db 1661 ACGGCAAGGCGCGCGCCATCGAATTCGCCCGCGCGAGAAACCGAGTCGTATATGCG 1720
: : : : :
QY 508 -----AlaProLeuTyLeuValLysLeu-----LysProLeuMetP 520
: : : : :
Db 1721 TGACCATCTTCCGACGCGCGCGCTACCGGTCAAGACCAAGAGAGATCTCTCA 1780
: : : : :
QY 520 heGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhe----- 535
: : : : :
Db 1781 TCGAGCGCGCGCGCGCTACCGCTGAGCGCGCGCGCTTCAGCGCGCGCGCGG 1840
: : : : :
QY 536 -----PheGluIleLeuIleProSerProThrSerThrSerProSerValProP 552
: : : : :
Db 1841 AGAAGTTCGAGTTCAGTATGCGCAGCAACAGCCCGCTCAAGCGCGAGC----- 1893
: : : : :
QY 552 roValValSerLysGlnProGlyAlaValGluValIleGlnTrpLeuThrMetGlyG 572
: : : : :
Db 1894 -----AGTACCGAGTGTCCCGCTGGAGG 1918
QY 572 InHiser-----LeuValGlyArgGlnTrpArgAlaPhePheAlaLysA 587
: : : : :
Db 1919 ACGCGCCACGACGAGCGGTCCGTCGCGCGCGCTTCGCTGAGCTTCGAGCGCGC 1978
: : : : :
QY 587 spAlaGlyTyArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProI 607
: : : : :
Db 1979 CCGGTACCTATTCCGTGAGCTGCGC-----GACCAAGGCGCGCAGC 2020
: : : : :
QY 607 leIleLysGluArgValHisPheAlaGluThrGlyIleThrPheArgProAsp----- 625
: : : : :
Db 2021 TCGTGGCGCCACCGCCACTCGTGGCGGCGGTGAAGTCGTGCGCGCACCG 2080
: : : : :
QY 626 -----ValPhe-----LysThrArgSerValValP 634
: : : : :
Db 2081 TCGAGTGTGTTCGACACCGCCGAGTACCGCACCGCGAGGAAGCCCTCCGACTGATCA 2140
: : : : :
QY 634 roAlaGluGluProValGlu-----GlnArgThrGluPheLysValS 648
: : : : :
Db 2141 CTTTCCCGCGCGGTGGAGATGCGCTTGTCTCTGGAGCGCGCAAGGTCGAGGCCA 2200
: : : : :
QY 648 erGlnMetLeu-----AspTrpLeu-----LeuGlnLeuAspAsnAnThr----- 661
: : : : :
Db 2201 CCGCCCTGTCTCAAGGCGCGCGACTGCTGCTGCGCTGGAATACTCAATCCGACCCAGT 2260
: : : : :
QY 662 -----TrpGlnProHisLeuLysLeuPheSer----- 670
: : : : :
Db 2261 ACGCGGTGTGATCCCGGTGCGCGAGGATTCCTACCGAACCTGACCTCTCGGTGCTCT 2320
: : : : :
QY 671 -----ArgIleGlnLeuGlyLeuSerLysT 679
: : : : :
Db 2321 ACACCAAGGCGCGCGACTACAGTTCACAGACGCGGGATCAAGGTGCGCATGCCCGAGG 2380
: : : : :
QY 679 hrTyAlaIleMetThrLeuGluProHisGlnIleArgHisLysThr----- 695
: : : : :

Db 2381 TAGAGTCGACATGCCACCCAGAGAGCGCTACGAAACCGGGGAGACGCGTGCACCGTGA 2440
QY 696 -----AspLeuLeuSerProSerGlyThrGlyValMetA 708
Db 2441 CCCTGGCCACGGCTTCGGCGCAACCGGTTTCAGCCACCTGACCTGACGCTGGTG 2500
QY 708 snAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyL 728
Db 2501 ATGAGATGCTATCCCTCGAGGCGAGATCGGCC--CGGCATCGACCATCTCTT-- 2555
QY 728 euGlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValI 748
Db 2556 -----CTACCATCCGCGCGCAACACGTGC--GCACGACG-- 2590
QY 748 leAspValAspThrGlyAspGluAspTrpIleGluThrTrpProSerGln---ArgL 767
Db 2591 -----CCAGCCTGGCTTCATCAGCTACGA 2615
QY 767 ystTrpGluCys-----AspPheValAspLysHisGlnArgThrL 780
Db 2616 CQTGCGTTGCCCGCAGACACGACGCGACCGCGGACGG--GCGAACCGGAGCGCGGCG 2674
QY 780 euGluVal-----ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnL 798
Db 2675 TCAAGGTACTGGAGCGCGCGCGAGGATGTCGACACCGCGCTGGCGAGCGGAGC 2734
QY 798 euLeuProValLeuGluAspArgAlaArgAspLysValLysMet-----A 813
Db 2735 TGGTCACGACGCGCCAGCGCAAGCGCTTCAGCTTCGCGATCCGACCTCGCTGACCC 2794
QY 813 rgLAlaIleGlyAspArgLeuIleAsnAsp-----LeuGlnArgGlnP 828
Db 2795 GCTGCGGATCACCGCGCGGCCATCGACGACACCGCGCAGTCCGCGCAGAGAGCAGT 2854
QY 828 heSerGluGlnLysHisAlaLeu-----AsnArgProValGluPheArgGln- 843
Db 2855 TCCTTCGCTCGAGAGAAGCGCTGTACCTGAAGTGGAGCGGCGCAGCTCGCTTCGCGCAGG 2914
QY 844 -----TrpValTyrGluSerTyrSerArgAlaThrArgV 856
Db 2915 GTGACACCGCGACTCGCTGTTCGCTTC-- 2946
QY 856 alSerHisGlyArgValPro-----PheLeuAlaGlyLeuProAspSerGln- 871
Db 2947 --AACGAGGCGAACAGCGCGG--CAAGCGCGAGCTGCTCAGCGCGCGCTGGCAGCAGC 3004
QY 872 --GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysGlnLysTyrL 891
Db 3005 CAGCCACACCTGGAGCTGGCCAAAGGAGTCAACTACATTCCTGCGCGCAACAGCGCGC 3064
QY 891 euGluAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuA 911
Db 3065 TCAGCATGGCGACTGGAGCGCGGAGCTGCGCAGGCGGCGCAGTCCGCGACCGCTGG 3124
QY 911 snIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuG 931
Db 3125 CGGTGCGCTTCAAC-----CTGTCGCGGACGCGCTGGCAGTGGAGC 3166
QY 931 luGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheT 951
Db 3167 AG----- 3168
QY 951 hrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspI 971
Db 3169 -----GTGCAGAACCTACGCTGGCGGCGGAGCAACCCGCTCAACTGCGCGCGGAGC 3223
QY 971 leGlnArgValArg-----AlaValPheLysProGluLeuH 983
Db 3224 CCGCGGAGTACGCTCGCTGGCGGATGGCGGCGCGGCGCTAC-----CTGG 3274
QY 983 isSerLeuLysAspValIlelePheSerThrLysGlyAsp----- 996
Db 3275 GCAATCTCGATGATCTGCTGAGTATCCTACGCGCGGCTCGAGCAGACCGCCAGCCAAC 3334

RESULT 4

US-08-258-261B-6

; Sequence 6, Application US/08258261B

; Patent No. 563949

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James M.

; APPLICANT: Beck, James Joseph

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Gaffney, Thomas Deane

; APPLICANT: Lam, Stephen Ting

; APPLICANT: Hammer, Phillip E.

; APPLICANT: Uknes, Scott Joseph

; TITLE OF INVENTION: Genes for the synthesis of

; TITLE OF INVENTION: antipathogenic substances

QY 997 --ValProLeuAla-----LysLysLeuSerGlyGlyAsp----- 1007
Db 3335 TGCTGCGCTGAGCATCGCTATCCGGCGCTGGCGGCGCGCGCGGATTCGGACCC 3394
QY 1008 -----TyrAspGlyAspMetA 1013
Db 3395 GCCTGGCGCTGATCATGCAAAACAGTCGCTGCGCTGTCGATGGCGCGCGGACG 3454
QY 1013 laTrpValCysTrpAspProGluIleValAspGly-----PheValAsnAla- 1028
Db 3455 CTTGGTTCCCTGGTGGGTGGCGAGTCGATGGGACGCTTTCCTCACCAGCTAGCCT 3514
QY 1029 -----GluMetProLeuGluProAspL 1036
Db 3515 ACTACGCCGACTGTACGCGCAGCGGCTGAGATCCAACTGCGCGCGGAA- 3567
QY 1036 euSerArgTyrLeuLysLysAspLysThrPheLysGlnLeuMetAlaSerHisGlyT 1056
Db 3568 -----CACTGGCAGCGCATCTCTCGAGCCCTAC- 3594
QY 1056 hrGlySerAlaLysGluGlnThrTyrAspMetIleGlnLysSerPheHisPheA 1076
Db 3595 -----GCCAAGCAGGCGAGCAGACCCCGCTGCTGCGAGCTGGCTGATCCTGG 3643
QY 1076 laLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleA 1096
Db 3644 CTTTC----- 3648
QY 1096 snAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGlyAsnLeuValA 1116
Db 3649 -----GCCCGCAGCATGCTGCTGCGGTGAACACCTGCTCGCGCGCTGCTCA 3697
QY 1116 spGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgGluL 1136
Db 3698 A-CGATCTCGCAAA-----CGCGCGCAAA 3720
QY 1136 euLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyA 1156
Db 3721 GGGCAGGCGCGCGGAGC-----CCGCTG--GAGCGCGCAGCGCTGGTGC 3768
QY 1156 rgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleA 1176
Db 3769 TGGCGCACCGCAGCTGCGCGCTGCGCGCTGCGCGCGCGAGTCTGCGCGTGGACCTTG 3828
QY 1176 spLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAsp----- 1191
Db 3829 CACGGCAACTGGGGTGGGTGGCGGTGCGCGCTGCGCGCGCGCGCGCGCGGAGCGGACGC 3888
QY 1192 -----ThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTrp 1209
Db 3889 AGCGCTGCTGCTGAGGCGCGCTTGGCGTTTCCCGATGCTTGTGCGCTCT- 3938
QY 1209 hePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu 1223
Db 3939 -----CGCTCGCGCGTGGATGGCGCAGCGCGCGCTCTGCTG 3977

6932 CCTCGCCATCCCGCCAGCCATGACCTGGGGCTTGGCGCGCTCATCGGCGCTCGAGC 6991
473 ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer----- 487
6992 ACCCGCAACGGTGGGAGGTCTCGTCGACGTCTCGCTGGGGTTCACAGAGCGCCGTGG 7051
488 -----LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
7052 GCCGTTCTGCTGCCGCCCTCGCCGAGCCCGACGAGAGACAGCTGCTCTCCGCCCGG 7111
502 ProAspAsn-----SerProThrAlaProLeuTyrLeuValLysLeuProLeu 518
7112 CGGAGCTCTAGCTCGCGCGCTGTCGCGCGCCCGCTCGCGGATCGGCTCCCG----- 7165
519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPheGluLe 538
7166 -----CGCGGACTTCACGC----- 7180
539 LeuIleProSerProThrSerProSerValPro-----ProValValSerLysGln 557
7181 -----CCGAGGACCACTTCATCATCCGCGCGGACCGCGCCATTGGCGCTCACGTCG 7234
558 Pro-----GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
7235 CCCGATGGCTCGCTCGAGAGCGCTCAGACCTCGCTCCTCATCA----- 7279
577 GlyArgGlnTrpArgAlaPheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
7279 ----- 7279
597 PheGlnLeuArgAlaGluAspProLysProIleLeuLysGluArgValHisPheAla 616
7280 -----GCGCGCGAGCGCGCGAGGCGCCCTG----- 7303
617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
7304 -----GCGCTCGAGCTCCACGAGAGCTCTCGGCCCTCGCGCGC 7345
637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
7346 GCACCACTCGCGCGCTCGCATGTCGCGCGACCGGATGTCGTCGCCACGCTTCTTGGC 7405
657 LeuAspAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeu 674
7406 ACCTCGAGCGCAAGGTCGAGGTCGCGCGGTTCCTCCAGCGAGCGCATCGAACACC 7465
675 GlyLeuSerLysThrTyAlaIleMetThrLeuGluProHis----- 688
7466 AGCTCCGCTCGACGCCACCTCTTTCCAGGATCTCGCGAGGTGTCTCGCGCAAGTCG 7525
689 GlnIleArgHisLysThrAspLeuLeuSer-----ProSerGlyThrGlyGluVal 706
7526 AAGGTGCAAG-CACCTCCAGACCTGCTCGGCTCTCGACCCCTCGACGCGCTTTGTCTC 7584
707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg----- 723
7585 TTTTCGTCGGGCGC-GGCGGCTGGGCGCGGACAGCGAGCGGCTACGCGCGCAAA 7643
724 -----AspValLeu 726
7644 GCGCTTCTCGACGCCCTTGGCGAGCATCGCGCAGCGCTGGATGACGAGCGTGGT 7703
727 GlyLeuGlyAspVal-----Pro 732
7704 GCGCTGGGCGGCTGGGCGCGGCGGATGGCCACCGATCAGCGCGCGCACCTCCA 7763
733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----TrpValIle 748
7764 ACAGCGCGGTCTGTGCGGATGGGCGGCTGCTGCGCTGCGCGCGCTGCGGTCT 7823
748 eAspValAspAspThr-----GlyAspGluAspTrpIleGluThrTyrProSe 764
7824 GGAGCAGCAGGACACCGCTCACCGTCCGCGACATCATGACTGGCGCGCTTGGCGCTTC 7883

764 r-----GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
7884 GTTACGCGCGCTCGCCCCCGCGCTCTGCGCGAT-----TTCCCGAGCGCGAGCGCGC 7940
779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
7941 TCTCGAGACCGAGCGCGCTCTCCGAGCATGGCCCGGC----- 7983
799 uProValLeuGluAspArgAlaAspLysValLysMetArgGlnAlaIleGlyAspAr 819
7984 -CCGACCTCTCTCGACAGCTCGGAGCGCTCGGAGCGAGCAGCTT-----CG 8033
819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
8034 TCTGCTCTGCTGCTGCTGGTGGCGACGAGAGCGCCCTCGCTCGCGCCAGAGCGCGCTC 8093
839 -----ValGluPheArgGlnTrp----- 844
8094 CCATGTGAGCGCGCGACAGGGCTTCTCGATCTCGGTCTCGATTCTGCTCATGGCGCTCGA 8153
844 ----- 844
8154 GCTTCCCGCGCGCTTGCACACAGCGCGCATCAAGCTCCGCGCACCTCGCTTCTCGA 8213
845 -----ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
8214 CCATCCCTCTCTCATCGAGTCGCGCTCTTCTTGGCGACTCGC-----TCGCCCA 8264
861 aProPheLeuAlaGlyLeuProAspSer 870
8265 CGCCCTCGGACGAGGCTCTCGCTCGAGC 8293
RESULT 5
US-08-456-837-6
Sequence 6, Application US/08456837
Patent No. 5643774
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,837
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-456-837-6

Alignment Scores:
Pred. No.: 0.000564 Length: 28958
Score: 178.50 Matches: 213
Percent Similarity: 29.87% Conserved: 89
Best Local Similarity: 21.07% Mismatches: 399
Query Match: 2.41% Indels: 312
DB: 1 Gaps: 42

US-09-913-878A-2 (1-1402) x US-08-456-837-6 (1-28958)

| | | | |
|----|------|---|---------------------------|
| QY | 38 | ProHisArgAlaGlys----- | -----GluLeuAlaGluSer 47 |
| DB | 5663 | CTCCGAGGAGGAGTCCCGACCTTCTCCAGCCCTAGAGGCCGAGCTAGCC---TCG | 5719 |
| QY | 48 | AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg | 67 |
| DB | 5720 | CGGCCCTCAATGGGCTCTCTCCACCGCTGCTCGTGGGATGAAGCGGTGTGGAGA | 5779 |
| QY | 68 | LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheLeuAlaLysAlaLaser | 87 |
| DB | 5780 | TCGCCGCCGAGCGGAGCCCTCGAGGAAAGACACACACCGCTCGCGGTGACCGCCT | 5839 |
| QY | 88 | Ser-----AsnTrpVal----- | -----ProLysAla--- 94 |
| DB | 5840 | TCCATTCCCGCAGATGGAGGAAATGCTCGACGACTTCGCGCGCTGCGCCAGACCTCA | 5899 |
| QY | 95 | -----HisLaspProAspThr----- | -----LeuProTrpSer 104 |
| DB | 5900 | CTTACCATCCGACCATCCCATCATCTCCAAAGTCCACCGCGCGCGCGCCGACGACC | 5959 |
| QY | 105 | LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnTrpValLeuLeu | 124 |
| DB | 5960 | ACGAGCTCGCTCGCCCGCTACTGGTCCGCCACGCTTCGCCACACCGCTCCGCTTCG | 6019 |
| QY | 125 | GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg | 144 |
| DB | 6020 | ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGACAGTGTCTTTCTCGAGCTCGGGCCTC | 6076 |
| QY | 145 | ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp | 164 |
| DB | 6077 | ACCTGTCTCTCGGCCCTTGGCAGACGCTCTGGACAGACGAGGAGGAGGAGGAGGAGG | 6136 |
| QY | 165 | GluProAla-----AsnValThrPheAlaAspProProLysArgSerLeuThrArg | 181 |
| DB | 6137 | GCGCCCTCTCCATCCCATCCCTCCGCAAGGAGCGGACGACGCGCGAGGCGTTCACCGCGGC | 6196 |
| QY | 182 | SerAlaThrGlyProProLysGlyAlaAlaThrPheLeuLysPheProAspProVal | 201 |
| DB | 6197 | TCGGCGCTCTCCACTCCG----- | -----CAGCATCACCCGACT 6232 |
| QY | 202 | AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg | 221 |
| DB | 6233 | GGAGCGTCTTCTCGCCCTTCGCTCCAC----- | ----- 6262 |
| QY | 222 | AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile | 240 |
| DB | 6263 | -----GCAAGTCTCTCCCTCCCACTTATGCTCTCCAGC | 6295 |

| | | | |
|----|------|---|--|
| QY | 241 | AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla | 260 |
| DB | 6296 | GCAGCGCTTCTGGCCGACGCTCCAGAGCACCGGCG-----CCGACGTCA | 6343 |
| QY | 261 | ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr | 280 |
| DB | 6344 | GCACCTTCTGCTCGCTCGAGGGGGGCTCTGCAAGCCATTCGAGCGCGGACCTCGATG | 6403 |
| QY | 281 | GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu | 300 |
| DB | 6404 | CGCTACGCGTCACTCCACGTGGACGCGCGCGCGCGCTCGCCCTCTCTCTCTCTCT | 6463 |
| QY | 301 | AlaProProSer----- | -----GlnProArgGluLysArgProValAsp 313 |
| DB | 6464 | TTCCCACTCTTCGAGCTTTCGCCACGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG | 6523 |
| QY | 314 | AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrLys | 333 |
| DB | 6524 | GCTACCGTA----- | -----TCACCTGGAAGCTCTGACACACCGCG 6559 |
| QY | 334 | SerHisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThr | 353 |
| DB | 6560 | AAACACCCCGCGACCTCGCGCGCACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG | 6619 |
| QY | 354 | TyrTyrGluSerPheProSerSerGlyGlyGluGlyValaIleProGluProSerArgSer | 373 |
| DB | 6620 | ---ACGGCTCTCTCGCGCTCA----- | -----CCGAGGCGCTCA 6652 |
| QY | 374 | AsnGlyLeuAlaArgSerGluGlu----- | -----SerAlaArg 384 |
| DB | 6653 | CCCGCGCGCGCGCGCGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 6712 |
| QY | 385 | SerGlnValGlnValHisAlaProValAlaAlaArgLeuArgAsnIleTrpProLys | 404 |
| DB | 6713 | CTC----- | -----TCGCCGACATCTCGCGCGAGCTTTCGCCGAGAGACGCCCGCA 6757 |
| QY | 405 | PheProLysTrpLeuHisGluAlaProLeuAlaValaIleTrpGluValThrArgLeuPhe | 424 |
| DB | 6758 | TTCCGCGCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCG | 6811 |
| QY | 425 | MetHisCysLysValaAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp | 444 |
| DB | 6812 | CTGCCCTCGCGCGCGCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCT | 6871 |
| QY | 445 | SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg | 464 |
| DB | 6872 | ACCTCGAGGCGCGCTTGTGTTCTTTCACGCGCGCGCGCTCTCTCTCTCTCTCTCTCT | 6931 |
| QY | 465 | GlyLysProPheProGluLysPro----- | ----- 472 |
| DB | 6932 | CCCTCGCCCATCCCGCCGAGGCTTACCTGGGCTTGGCGCGCTCATCGGCTCGAGC | 6991 |
| QY | 473 | -----ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer----- | 487 |
| DB | 6992 | ACCCGACCGCTGGGGAGGTCTCGTCGACGCTCTGGGTGGGTGCGAGAGGCGCGCTGG | 7051 |
| QY | 488 | -----LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn | 501 |
| DB | 7052 | GCCTGCTGCTCGCGCGCTCGCGAGCGGCCCGACAGACAGCTCTCTCTCGGCGCG | 7111 |
| QY | 502 | ProAspAsn-----SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu | 518 |
| DB | 7112 | CCGAGCTCTACGCTCGCGCGCTCTCGTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCT | 7165 |
| QY | 519 | MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPheGluIle | 538 |
| DB | 7166 | -----CGCGGACTTTCACG----- | ----- 7180 |
| QY | 539 | LeuIleProSerProThrSerProSerValPro---ProValValSerLysGln | 557 |
| DB | 7181 | -----CCGAGGCGACATTCTCATCCCGCGCGCGCGCTTGGCGCTCACGCTCG | 7234 |

| | | | | |
|----|------|--|---|------|
| QY | 558 | Pro | ---GlyAlaValGluValleGlnTrpLeuThrMetGlyGlnHisSerLeuVal | 576 |
| Db | 7235 | CCC | ATGCGTCCGCTCGAAGCGCCCTCAGACCTCGCTCATCA | 7279 |
| QY | 577 | GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu | 596 | |
| Db | 7279 | --- | --- | 7279 |
| QY | 597 | PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla | 616 | |
| Db | 7280 | --- | ---GCGCGGAGGCGCCAGGCGCCGCT | 7303 |
| QY | 617 | GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu | 636 | |
| Db | 7304 | --- | ---GCGCTCCGAGCTCCAGCAGCTCTCGGCCCTCGCGCGC | 7345 |
| QY | 637 | GluProValGlnArgThrGluPheLysValSerGlnMetLeuAspTyrLeuLeuGln | 656 | |
| Db | 7346 | GCACCA | CCCTCGCGCGCTGCATGTCGCCAGCGAATGCTGCGCACGCTCTTGGAGC | 7405 |
| QY | 657 | LeuAspAsnAnThrTrpGlnProHisLeuLysLeuPheSerArgIle | 674 | |
| Db | 7406 | AGCTCGAGCGCGAAGGTCGACGAGTCCGCGCGGTTCACGCGAGCGGCATCGAACACC | 7465 | |
| QY | 675 | GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis | 688 | |
| Db | 7466 | ACGCTCGCTCGAGCCACCTCTTCAGGATCTCGCGAGGTGTCTCGGCAAGTCG | 7525 | |
| QY | 689 | GlnIleArgHisLysThrAspLeuLeuSer | 706 | |
| Db | 7526 | AAAGTGAAG-CACCTCCAGCACTCTCGGCTCTCGACCCCTCGACGCTTTGTTCTC | 7584 | |
| QY | 707 | MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg | 723 | |
| Db | 7585 | TTTTCGTCCGCGC-GGCGCTCTGGCGCGGCGACAGAGCGCTTACGCGCGCGCAA | 7643 | |
| QY | 724 | --- | ---AspValLeu | 726 |
| Db | 7644 | CGCTTCTCTCGAGCGCCCTTCGCGAGCATCGCGCAGCGCTGGATTGACAGCGAGTGGT | 7703 | |
| QY | 727 | GlyLeuGlyAspVal | 732 | |
| Db | 7704 | GGCTGGGCGGTGGGGCGGCGCGCATGCGCACCGATCAGCGGCGACGCCACCTCCA | 7763 | |
| QY | 733 | SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet | 748 | |
| Db | 7764 | ACAGCGGGTCTGTGCGGGATGGCCCTCGCTTGCCCTGCGCGGCTCGCGTGGCTCT | 7823 | |
| QY | 748 | eAspValAspThr | 764 | |
| Db | 7824 | GGAGCAGCAGACACCACTCACCGTCGCGGCATCATGCTGCGCGCTTTTCGCTTC | 7883 | |
| QY | 764 | r | ---GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThr | 779 |
| Db | 7884 | GTTCAGCGCGCTCGCCCGCCCGCTCTCTCGCGCAT | 7940 | |
| QY | 779 | rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuLeuLe | 799 | |
| Db | 7941 | TCTCGAACACCGAAGCGCGCTCTCCGAGCATGGCCCGCC | 7983 | |
| QY | 799 | uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr | 819 | |
| Db | 7984 | -CCGACCTCTCGACAAGTCTCGGACCCCTCGGAGCGGACGACTT | 8033 | |
| QY | 819 | gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro | 838 | |
| Db | 8034 | TCTGCTCTCTGCTGGTGGCCACGAGACGCGCTCTGCTCTCGATCGCTCATGCGCTCGA | 8153 | |
| QY | 839 | --- | ---ValGluPheArgGlnTrp | 844 |
| Db | 8094 | CCATGTCGACCCCGACAAAGGGCTTCTCTGATCTCGGTTCTGATCGCTCATGCGCTCGA | 8153 | |
| QY | 844 | --- | --- | 844 |

Bd 8154 GCTTCGCCGGCGGTTGCACAGCCACCAGCATCAAGTCCC CGGCCAACCCCTC GCCTTCGA 82121

Qy 845 -----ValTyrgLuseTy-SerSerArgAlathrArGValseHisGLyArgV 861
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dd 8214 CCATCCCTCTCCCTCATCGACTCGCGGTCTTCTTGCGGACTCGC-----TCGCCCA 8264

Qy 861 alProPheLeulaGLyleuProAspSer 870
|||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

Dd 8265 CCGCCTCGGCACGAGGCTCTCCGTCGAGC 8293

RESULT 6
US-08-457-342--6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TYPE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-457-342--6

Alignment Scores:

| Pred.No.: | Score: |
|------------------------|--------|
| 0.000564 | 178.50 |
| Percent Similarity: | 29.87% |
| Best Local Similarity: | 21.07% |
| Query Match: | 2.41% |
| Mismatches: | 89 |
| Indels: | 332 |

Length: 28958
Matches: 213
Conservative: 89
Mismatches: 339
Indels: 312

DB: 1 Gaps: 42
US-09-913-878A-2 (1-1402) x US-09-457-342-6 (1-28958)
QY 38 ProHisArgArgLys-----GluLeuAlaGluSer 47
Db 5663 CTTCCGAGAGGAGTCCGCGACCTTCCAGCCCTACAGAGCCGAGCTAGCC---TCG 5719
QY 48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db 5720 CGCCCTCAATGGGCTCTCTCCACCGCTGCTGCGGATGAAGACGCGGTGGAGA 5779
QY 68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
Db 5780 TCGCCCGCCAGCCGAGCCCTCGAGCAAGAACACACAGCGCTGCGGCTCAGCCAGCCT 5839
QY 89 Ser-----AsnTrpVal-----ProLysAla--- 94
Db 5840 TCCATTCCCGCAGCATGGACGAAATGCTCGACGACTTCGCGCGCTCGCCAGAGCTCA 5899
QY 95 -----HisAlaAspProAspThr-----LeuProTrpSer 104
Db 5900 CCTACATCCCGACGATCCCATCATCTCCAGCTCAGCGGCGGCGGCCAGGACC 5959
QY 105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnTrpValLeuLeu 124
Db 5960 ACGAGCTCGCTCGCCGACTACTGGTCCGCGACGTTCCGACACCGCTCGCTTCCTCG 6019
QY 125 GluValLeuAsnArgPheMetProProAsnAsnThrProGlyArgTrpPheGlyArg 144
Db 6020 ACGCGGTAGTGGCTTC---ACGCCAAGGGGCACGTTCTTCTCGAGCTCGGCGCTC 6076
QY 145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db 6077 ACGTGTCTCTCCGCTTCGCGCAAGACGCGCTCGGACAGGACGAGGACGCTCGCAT 6136
QY 165 GluProAla-----AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db 6137 GCGCTCTCTCCACCTCCGCAAGGACGCGACGCGGAGCGGTACCGCGCGGC 6196
QY 182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db 6197 TCGCGCTCTCCACTCCG-----CAGGCATCACACCGACT 6232
QY 202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db 6233 GGAGCGCTTCTTCGCGCCCTTCGCTCCAC----- 6262
QY 222 AlaLysGlyLysLeuSerAsnValAlaAlaAlaAlaProProValPro---Ile 240
Db 6263 -----GCAAGTCTCCCTCCCACTATGCTTCAGC 6295
QY 241 AlaSerAlaLeuAspLysValProThrArgHisAlaAsnThrArgAspProThrAla 260
Db 6296 GCGAGCGCTTCTGCGCGCAGCGCTCCAGGACCGCGG-----CCGACGTCA 6343
QY 261 ThrGlyHisArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db 6344 GCCACCTTCGCTCGCTCGAGGGGGGCTTCGCAAGCCATCAGCGCGGAGCTTCGATG 6403
QY 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db 6404 CGCTCAGCGGTACGTCCACGTGAGCGGCGGCGCGCGCGCTCGCTCGCTTC 6463
QY 301 AlaProProSer-----GlnProArgGluLysArgProValAsp 313
Db 6464 TTCCCACTCTCGAGCTTCGCCAGGAGCGGCAAGAGCAGCGGTCCAGCGCTGGC 6523
QY 314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db 6524 GCTACCGTA-----TCACCTGGAAGCTCTGACCAACCGCGC 6559
QY 334 SerHisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThr 353

Db 5660 AAACACCCGCGGACCTCGCGGACCTGGCTCGTCTGTCGGCGCGCTCTGGAGCAGC 5619
QY 354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProLysProSerArgSer 373
Db 5620 ---ACGCGCTCCCTCCGCGCTCA-----CCGAGGCGCTCA 6652
QY 374 AsnGlyLeuAlaArgSerGluGlu-----SerAlaArg 384
Db 6653 CCGCGCGCGCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 6712
QY 385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db 6713 CTC-----TCGCGGAGCATCTCGCGCAGGCTTCGCGCGGAGACCGCCCGCA 6757
QY 405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
Db 6758 TTCGCGCGGTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 6811
QY 425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLysTyrAspProSerTrp 444
Db 6812 CTGCGCTCGCGCGGACTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCG 6871
QY 445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db 6872 ACCTCAGGCGCTTGTGTTCTTTCACGCGCGCGCGCTCTCCATTGGACACTCTGACC 6931
QY 465 GlyLysProPheProLysPro----- 472
Db 6932 CCTCTCGCCCTCCGCGCGGACCATGACCTGGGCTTGGCGCGCTCATCGCGCTCGAC 6991
QY 473 -----ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer----- 487
Db 6992 ACCCGACCGGTGGGAGGTCTCGTCGACGTCTCGCTCGCTCGCTCGCTCGCTCGCT 7051
QY 488 -----LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db 7052 GCGCTTGTTCGCGCGCTCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7111
QY 502 ProAspAsn-----SerProThrAlaProLeuTyrLeuValLysLysLysProLeu 518
Db 7112 CCGGACTCTACGCTCGCGCATCTCGTCGCGCGCGCGCTCGCGCGCTCGCGCTCGCG 7165
QY 519 MetPheGluGlnClyCysArgLeuThrArgArgPheGlyProAspArgPheGluIle 538
Db 7166 -----CGCGGACTTCACGC----- 7180
QY 539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
Db 7181 -----CCGAGGACCATCTTCATCACCGCGGCGGCGGCGGCGGCGGCGGCGGCG 7234
QY 558 Pro---GlyAlaValGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
Db 7235 CCGGATGGCTCGCTCGAAGAGCGGCTCAGACCTCGCTCTCATCA----- 7279
QY 577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
Db 7279 ----- 7279
QY 597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
Db 7280 -----GCCCGGAGCGCGGAGCGCGCTG----- 7303
QY 617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
Db 7304 -----GCGCTCGGAGCTCCAGCAGGCTCTCGCGCTCGCGCGCGCGCGCG 7345
QY 637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuGln 656
Db 7346 GCACCACTCTCGCGGCTCGATGTCCGCGACCGGAATCTGTTCGCGACCGCTTCTGAGC 7405
QY 657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeu 674

```

Db 7406 AGCTGACGCCGAGAGGTGCGACGGTCCGCGCGCGTGTTCACCGGAGCGGCATCGAACACC 7465
Qy 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis----- 688
Db 7466 AGCTCCGCTCGACGCGACCTCTTTTCAGGATCTCGCGAGGTGTCTCCGCGAAGTCG 7525
Qy 689 GlnIleArgHisLysThrAspLeuSer-----ProSerGlyThrGlyGluVal 706
Db 7526 AGGTGCAAG-CACCTCCACGACCTGCTGGCTCTCGACCCCTCGACGCGCTTTGTTCTC 7584
Qy 707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgileArg----- 723
Db 7585 TTTTCGTCGGCGC-GGCGGTCTGGGGCGCGGACAGCAGCGCGCTACGCGCGCGCAA 7643
Qy 724 -----AspValLeu 726
Db 7644 CGCTTCTCTCGACGCGCTTTCCGAGCATCGCGCAGCGTGGATTTGACAGCGACGCTCGT 7703
Qy 727 GlyLeuGlyAspVal-----Pro 732
Db 7704 GGCTGGGGCGGTGGGGCGGGCGGGGATGCGCCACCGATCAGGGCGGAGCGCCACCTCCA 7763
Qy 733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----TrpValI 748
Db 7764 ACAGCGCGTCTGTTCGGGATGGCCCTCGCTTGGCTTGGCGGCGCTCGCGCTGGCTCT 7823
Qy 748 eAspValAspThr-----GlyAspGluAspThrIleGluThrTyrProSe 764
Db 7824 GGAGCAGCAGACACCGTCACCGTCGCTCCGCCACATCGACTGGCGCGCTTTGCGCCCTC 7883
Qy 764 r-----GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
Db 7884 GTTCAGCGCGCTCGCGCGCGCGCTCTCTCGCGAT--TTGCGCGAGCGCGAGCGGC 7940
Qy 779 rleuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 799
Db 7941 TCTCGAGACAGCAGGAGCGCTCTCTCGAGCATGCGCGCGC----- 7983
Qy 799 uPrValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
Db 7984 -CCGACCTCTCGACAGCTCGAGCGCGCTCGGAGCGGAGCGAGCT-----CG 8033
Qy 819 gLeuIleAsnAspLeuGlnArgPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
Db 8034 TCTGCTCGTCTGCTGTGCGGACGAGAGCGCGCTCTCTCGCGCGCGCTCGCGCGCTC 8093
Qy 839 -----ValGluPheArgGlnTrp----- 844
Db 8094 CCATGTGACCCCGACAGAGGCTTCTCGATCTCGGTCTCGATTCTCATGGCGCGTCA 8153
Qy 844 ----- 844
Db 8154 GCTTCGCGCGCTTGCAACAGCGCACCGGCATCAAGCTCCGCGCACCTCGCGCTTCA 8213
Qy 845 -----ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
Db 8214 CCATCCCTCTCTCATCGAGTGGCGCTCTTCTTGGCGGACTGCG-----TCGCCCA 8264
Qy 861 aLProPheLeuAlaGlyLeuProAspSer 870
Db 8265 CGCCCTCGCACGAGGCTCTCGTGCAGC 8293

```

RESULT 7

```

US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane

```

```

; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip B.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-646A-6
;
Alignment Scores:
Pred. No.: 0.000564 Length: 28958
Score: 178.50 Matches: 213
Percent Similarity: 29.87% Conservative: 89
Best Local Similarity: 21.07% Mismatches: 399
Query Match: 2.41% Indels: 312
DB: 1 Gaps: 42
;
US-09-913-878A-2 (1-1402) x US-08-457-646A-6 (1-28958)
Qy 38 ProHisArgArgLys-----GluLeuAlaGluSer 47
Db 5663 CCTCCGAGGAGGAGTCCGCGACCTTCTCCAGCCCTACGAGCCGAGCTAGCC---TCG 5719
Qy 48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db 5720 CGCCCTCATGGCGCTCTCTCCACCGTCTGCTCGCATGAGAGCGCGGTGGAGA 5779
Qy 68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaLysSer 87
Db 5780 TCGCCCGCCGAGCGCGAGCGCTTCGAGAAAGACACACGCTTCGCGCTCAGCCGCT 5839
Qy 88 Ser-----AsnTrpVal-----ProLysAla--- 94
Db 5840 TCCATTCCCGCACATGGACGGAAATGCTCGACGACTTCGCGCGCTCGCCGAGGCTCA 5899
Qy 95 -----HisAlaAspProAspThr-----LeuProTrpSer 104

```

Db 5900 CCTACCATCCGACGATCCCATCATCTCCAAAGTCCACCGCGCGCGCCACGAGC 5959
 Qy 105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
 Db 5960 ACAGCTCCGCTCCGCGGACTACTGGTCCGACAGTTCGCACACACCTCCGCTCCCTCG 6019
 Qy 125 GluValLeuAsnArgPheMetProProAsnAsnThrProGlyArgThrPheGlyArg 144
 Db 6020 ACGCGGTACGTGCCCTTC---ACGCGAAGGGGCACGTGTCTTTCTCGAGTCCGGCCTC 6076
 Qy 145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
 Db 6077 ACCTGTCTCTCCGCGCTTGGCGAAGACGCCCTCGACAGGACGAGCGACGTCCGCAT 6136
 Qy 165 GluProAla-----AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
 Db 6137 GCGCCCTTCCTCCACCCCTCCGCAAGGACGCGAGCGCGAGCGGTTCACCGCGCGC 6196
 Qy 182 SerAlaThrGlyProProLeuHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
 Db 6197 TCGCGCTCTCCACTCCG-----CAGCGATCACCCCGACT 6232
 Qy 202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
 Db 6233 GGAGCGCTTCTTCGCGCCCTTGCTCCAC----- 6262
 Qy 222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
 Db 6263 ---GCAAGTCTCTCCCTCCCTCCACCTATGCTTCCAGC 6295
 Qy 241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
 Db 6296 GCGAGCGCTTCTGGCGCGAGCTCCCAAGGACCCCGCG---CCGACGTCA 6343
 Qy 261 ThrGlyHisArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
 Db 6344 GCCACCTTGTCTCGCTCGAGGGGGGCTCTGGCAAGCCATCAGCGCGGGACCTCGATG 6403
 Qy 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
 Db 6404 CGTACGGGTGAGTCCACGTGAGCGGACGAGCGGGCGCGCGCTCGCTCGCTGCC 6463
 Qy 301 AlaProProSer-----GlnProArgGluLysArgProValAsp 313
 Db 6464 TTCCACCCCTCTCGAGCTTTCGCGACGAGCGGCAAGACGAGACGCGTCGACGCTGCG 6523
 Qy 314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrLys 333
 Db 6524 GCTACCGTA-----TCACCTGGAGGCTCTGACACCGCG 6559
 Qy 334 SerHisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThr 353
 Db 6560 AAACACCGCGGACCTCGCGGACCTGCTCGTCTCGCGCGCGCTCTGGACGAG 6619
 Qy 354 TyrTyrGluSerPheProSerSerGlyGlyGlyAlaIleProGluProSerArgSer 373
 Db 6620 ---ACGCGCTCCCTCGCGCTCA-----CCGAGGCGCTCA 6652
 Qy 374 AsnGlyLeuAlaArgSerGluGlu-----SerAlaArg 384
 Db 6653 CCGCGCGCGCGCGGCTCTCGCTTGGCTGCGCTGAGCCAGGCGCCACCTGGACCGGAG 6712
 Qy 385 SerGlnValGlnValHisAlaProValAlaAlaArgLeuArgAsnIleTrpProLys 404
 Db 6713 CTC-----TCGCGAGCATCTGCGCCAGGCTTCGCGGAGACCGCCCCGA 6757
 Qy 405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaIleTrpGluValThrArgLeuPhe 424
 Db 6758 TTGCGCGGTGCTCTCGCTCTCGCTCTCGACGAGCGCC-----CCCTCGACGACGCTC 6811
 Qy 425 MethCysLysValAspLeuGluAspGluSerLeuGlyLysThrAspProSerTrp 444
 Db 6812 CTGCGCTCGCGCGGACTCGCGCTCTCGCTTCTCTCAAGCCCTCGCGGACCTCG 6871

Qy 445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyArgLeuAspAlaPheArg 464
 Db 6872 ACCTCGAGCGCGCTTGTGTCTTTCACGCGCGCGCGCTTCTCATTTGACACTCTGACC 6931
 Qy 465 GlyLysProPheProGluLysPro----- 472
 Db 6932 CCCTCGCCCATCCCGCCAGGCATGACCTGGGGCTTGGCGCGCTCATCGGCTCGAGC 6991
 Qy 473 -----ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer----- 487
 Db 6992 ACCCGACCGGTGGGAGGTCTCGTCGAGCTCTCGCTGGGTGCGACGAGCGCGCTGG 7051
 Qy 488 -----LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
 Db 7052 GCGCTTGTCTGCGCGCGCTCGCGAGCGGCACGACGAGACCATGCTCTCCGCCCG 7111
 Qy 502 ProAspAsn-----SerProThrAlaProLeuTyLysValLysLeuLysProLeu 518
 Db 7112 CCGGACTCTAGCTCGCGCATCTCGCGCGCCCGCTCGCGCATGCGCTCCCG----- 7165
 Qy 519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPheGluIle 538
 Db 7166 -----CGCGCACTTCAACG----- 7180
 Qy 539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
 Db 7181 -----CCGAGGACCATTTCTCATCCGCGCGCACCGCGCATTTGGCGCTCACGTCG 7234
 Qy 558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
 Db 7235 CCCGATGCTCGCTCGAAGAGCGCTCAGCACCTCGCTCTCATCA----- 7279
 Qy 577 GlyArgGlnTrpArgAlaPheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
 Db 7279 ----- 7279
 Qy 597 PheGlnLeuArgAlaGluAspProLysProIleLysGluArgValHisPhePheAla 616
 Db 7280 -----CGCGCGAGGCGCGAGCGCTG----- 7303
 Qy 617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
 Db 7304 -----GCGCTCGAGCTCCACGACGAGCTCGCGCTCGCGCGC 7345
 Qy 637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
 Db 7346 GCACACCTCGCGCGCTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405
 Qy 657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeu 674
 Db 7406 AGCTCGACCGCGAAGGTGCGAGGTCCGCGCGCTTCCAGCGGCGGCGATCGAACACC 7465
 Qy 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis----- 688
 Db 7466 ACGCTCGCTCGACCGCACCTCTTTCAGGATCTCGCGAGGTGTCTCGGCAAGGTG 7525
 Qy 689 GlnIleArgHisHisLysThrAspLeuLeuSer-----ProSerGlyThrGlyGluVal 706
 Db 7526 AAGGTGCAAG-CACCTCCACGACCTGCTCGCTCTCGACCCCTCGACGCTTGTGTCTC 7584
 Qy 707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg----- 723
 Db 7585 TTTTCGTCCGCGC-GGCCGTCTGGGCGCGGACGACGAGCGCTACGCGCGCGCAAA 7643
 Qy 724 -----AspValLeu 726
 Db 7644 CGCTTCTCGACCGCTTTCGCGAGCATCGCGCGAGCTGGATTGACGACGAGCTCGGT 7703
 Qy 727 GlyLeuGlyAspVal-----Pro 732
 Db 7704 GGCCTGGGCGCTGGGCGCGCGGCGATGCGCACCGATCAGCGCGGAGCCACCTTCCA 7763

QY 202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
 Db 6233 GGAGCGCTTCTCGCCCTTCGCTCCAC
 QY 222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProValPro---Ile 240
 Db 6263 -----CCAAGTCTCCCTCCACCTATGCTTCCAGC 6295
 QY 241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
 Db 6296 GCGAGCGCTTGGCCCGACGCTCCCAAGGCACCGCGG-----CCGACGTCA 6343
 QY 261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrThrSerGlnGlyThrSerTyr 280
 Db 6344 GCCACTTGTCTCGTGGGGGGGCTTGGCAAGCATCGAGCGGGGACCTCGATG 6403
 QY 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
 Db 6404 CGCTCAGCGGTGAGTCCACGTGAGCGGCGAGCGCGCGCTCGCTCCCTGCTCC 6463
 QY 301 AlaProProSer-----GlnProArgGluLysArgProValAsp 313
 Db 6464 TTCCACCTCTCGAGCTTTCGCCAGCGGCAAGACAGACAGCGTCCAGCCTGGC 6523
 QY 314 AlaThrValPheGluAlaGlyHisLeuLeuGluSerProSerLysGlyArgThrThriys 333
 Db 6524 GCTACCGTA-----TCACCTGGAAGCTCTGACCAACCGCG 6559
 QY 334 SerHisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThr 353
 Db 6560 AAACACCCCGACCTCGCGGACCTGGTCTGCTGTCGCGCGCTCTGACGACG 6619
 QY 354 TyrTyrGluSerPheProSerSerGlyGlyGluAlaIleProGluProSerArgSer 373
 Db 6620 ---ACGCGCTCCCTCGCGCTCA-----CCGAGCGCTCA 6652
 QY 374 AsnGlyLeuAlaArgSerGlu-----SerAlaArg 384
 Db 6653 CCGCGCGCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 6712
 QY 385 SerGlnValGlnValHisAlaProValAlaAlaArgLeuArgAsnIleTrpProLys 404
 Db 6713 CTC-----TCGCGGAGCATCTCGCGGAGCTTTCGCGGAGCGCGCGCG 6757
 QY 405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
 Db 6758 TTCGCGCGCTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 6811
 QY 425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLysTrpAspProSerTrp 444
 Db 6812 CTGCTGCGCGCGGACTCGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 6871
 QY 445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
 Db 6872 ACCTCAGGCGGCTTGTGTTTTCACGCGCGCGCGCTCTCTCATTTGACACTCTGACC 6931
 QY 465 GlyLysProPheProGluLysPro----- 472
 Db 6932 CCTCGCCCATCCGCGCGGCGCATACCTGGGGCTTGGCGCGCTCATCGCTCGAGC 6991
 QY 473 -----ProAsnAspValPheValThrAlaMetThrGlyAsnPheLys----- 487
 Db 6992 ACCCGGCGCTGGGAGGTCTGCTGACGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7051
 QY 488 -----LysGlySerAlaValValLeuSerAlaValValLeuAspTyrAsn 501
 Db 7052 GCGCTTGTGCGCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7111
 QY 502 ProAspAsn-----SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
 Db 7112 CCGGACTCTACGCTCGCGCATCGTCCGCGCGCGCTCGCGCGCTCGCGCGCTCGCGCT 7165
 QY 519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPheGluIle 538

7166 -----CGCGGACTTCACGC----- 7180
 QY 539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
 Db 7181 -----CCGAGCGCACTTCTCATCAGCGGCGCACCGCGGCGCTTGGCGCTCAGCTCG 7234
 QY 558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
 Db 7235 CCGATGGTCTCGTGAAGAGCGGCTCAGCACTCTGCTCTCATCA----- 7279
 QY 577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
 Db 7279 ----- 7279
 QY 597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
 Db 7280 -----GCCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7303
 QY 617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
 Db 7304 -----GCGCTCGGAGCTCCACGACGAGCTCTCGGCGGCTCGGCGGCGGCGGCGG 7345
 QY 637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
 Db 7346 GCACCACTCTCGCGCTCGGATGTCGCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405
 QY 657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeu 674
 Db 7406 AGCTCAGCGCGAGGCTCGCAGTCCGCGCGCTGTTCACGCGAGCGCATCGAACACC 7465
 QY 675 GlyLeuSerLysThrTyraIleMetThrLeuGluProHis----- 688
 Db 7466 ACCTCTCGCTCGACGCGCTCTCTTCAGGATCTCGCGAGTGTCTCGGCAAGGTCTG 7525
 QY 689 GlnIleArgHisLysThrAspLeuLeuSer-----ProSerGlyThrGlyGluVal 706
 Db 7526 AAGTCCAAAG-CACCTCCAGCACCTGCTCGGCTCTCGACCGCTTGTGTCT 7584
 QY 707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg----- 723
 Db 7585 TTTTCTCGCGCGC-GCGCTCTGGGCGCGGACGACGAGGCGGTCTGCGGCGCAAA 7643
 QY 724 -----AspValLeu 726
 Db 7644 CGCTTCTCTCGACGCTTTCGCGGATCGCGGCGCTGCTGATTCACAGCGACGCTCGGT 7703
 QY 727 GlyLeuGlyAspVal-----Pro 732
 Db 7704 GCGCTGCGGCGCTGCGGCGCGCGCGCATGCGCACCGATCAGCGCGGCGGCGGCGGCGG 7763
 QY 733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----TrpValIle 748
 Db 7764 ACAGCGGCTCTCGCGGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7823
 QY 748 eAspValAspAspThr-----GlyAspGluAspTrpIleGluThrTyrProSe 764
 Db 7824 GGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 7883
 QY 764 r-----GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
 Db 7884 GTTCAGCGCGCTCGCGCGCGCGCTCTCTCGCGAT---TTGCGCGGCGGCGGCGGCGG 7940
 QY 779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 799
 Db 7941 TCTGAGACCGAGGCGGCGGCTCTCTCGAGCATGCGCGCGGCGGCGGCGGCGGCGGCGG 7983
 QY 799 uProValLeuGluAspArgAlaAspLysValLysMetArgGlnAlaIleGlyAspAr 819
 Db 7984 -CCGACCTCTCGACAGCTCCGAGCGCTCGGAGCGGAGCGGAGCTT-----CG 8033
 QY 819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838

QY 314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
| | | | |
Db 6524 GCTACCGTA- - - - -TCACCTGGAAGCCTCTGACCAACCGCG 6559
| | | | |
QY 334 SerHisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThr 333
| | | | |
Db 6560 AAACACCCCGCGACCTCCCGCGACCTCGCTCGCTCGCGCGCTCTGACGACG 6619
| | | | |
QY 354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
| | | | |
Db 6620 - - -ACGCGCTCCCTCCGCGCTCA- - - - -CCGAGCGCGCTCA 6652
| | | | |
QY 374 AsnGlyLeuAlaArgSerGluGlu- - - - -SerAlaArg 384
| | | | |
Db 6653 CCCGCGCGCGCGCGCTCTCGCTTGGCGCTGAGCCAGCGCCACCTGACCGCGAGG 6712
| | | | |
QY 385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTyrProLys 404
| | | | |
Db 6713 CTC- - - - -TCGCGAGCATCTCGCGCAGGCTTGGCGGAGACCGCCCGA 6757
| | | | |
QY 405 PheProLysTyrPheHisGluAlaProLeuAlaValAlaTyrPheGluValThrArgLeuPhe 424
| | | | |
Db 6758 TTCGCGGCGTCTCTCGCTCTCGCTCTCGCGCTGACGAGCGCC- - - - -CCCTCGCAGACCGTC 6811
| | | | |
QY 425 MethHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTyr 444
| | | | |
Db 6812 CTGCGCTGCCCGCGACTCCGCTCTCGCTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871
| | | | |
QY 445 SerThrAlaArgAspValThrAspIleTyrPheLysThrLeuTyrArgLeuAspAlaPheArg 464
| | | | |
Db 6872 ACCTCGAGCGCGCTTGTGGTCTTTCACGCGCGCGCGCTCTCCATTGGACACTCTGACC 6931
| | | | |
QY 465 GlyLysProPheProGluLysPro- - - - - 472
| | | | |
Db 6932 CCCTCGCCCATCCCGCCGAGCCATGCTGGGCTTGGCGCGTCTCGGCTCGAGC 6991
| | | | |
QY 473 - - - - -ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer- - - - - 487
| | | | |
Db 6992 ACCCGACCGGTGGGAGGCTCTCGCGAGCTCTCGCTGGGCTCGAGAGCGCGCTGG 7051
| | | | |
QY 488 - - - - -LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
| | | | |
Db 7052 GCGGCTTGTGCGGCGCTCGCGAGCCGACGAGACGAGACGCTGCTCTCGCGCGG 7111
| | | | |
QY 502 ProAspAsn- - - - -SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
| | | | |
Db 7112 CCGGACTCTACGCTCGCGCGCATCGCTCGCGCGCGCTCGCGCGCTCGCGCTCCCG- - - - - 7165
| | | | |
QY 519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
| | | | |
Db 7166 - - - - -CGCGCGACTTCACGC- - - - - 7180
| | | | |
QY 539 LeuIleProSerProThrSerThrSerProSerValPro- - - - -ProValValSerLysGln 557
| | | | |
Db 7181 - - - - -CCGAGGCGACCATCTCAATCACCGGCGGCGACCGCGCATTTGGCGCTCACGTCG 7234
| | | | |
QY 558 Pro- - -GlyAlaValGluGluValIleGlnTyrPheThrMetGlyGlnHisSerLeuVal 576
| | | | |
Db 7235 CCCGATGCGCTCGAAGAGCGCTCAGCACCTCGCTCTCTCATCA- - - - - 7279
| | | | |
QY 577 GlyArgGlnTyrArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
| | | | |
Db 7279 - - - - - 7279
| | | | |
QY 597 PheGlnLeuArgAlaGluAspProLysProIleLeuLysGluArgValHisPhePheAla 616
| | | | |
Db 7280 - - - - -GCCGCGAGCGCGCGAGGCCCTG- - - - - 7303
| | | | |
QY 617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
| | | | |
Db 7304 - - - - -CGCGCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGCGCGC 7345
| | | | |

QY 637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTyrPheLeuGln 656
| | | | |
Db 7346 GCACACACCTCCCGCGCTCGATGTCCGCGAGCGGAATGCTGCCACGCTTCTTGAGC 7405
| | | | |
QY 657 LeuAspAsnAsnThrTyrTyrGlnProHisLeuLysLeuPheSerArgIle- - - - -GlnLeu 674
| | | | |
Db 7406 AGCTCGAGCGCGAAGGTGCGAGGTCCGCGCTTCCACGCGAGCGGATCGAACACC 7465
| | | | |
QY 675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis- - - - - 688
| | | | |
Db 7466 AGCTCCGCTCGACGCCACCTCTTTCAGGATCTCGCGAGGTGTCTCCGCAAGGTGCG 7525
| | | | |
QY 689 GlnIleArgHisLysThrAspLeuLeuSer- - - - -ProSerGlyThrGlyGluVal 706
| | | | |
Db 7526 AAGGTGCAAG-CACTCCACGACCTGTCTCGCTCTGACCCCTCGAGCCCTTTCTTC 7584
| | | | |
QY 707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg- - - - - 723
| | | | |
Db 7585 TTTTGCTCGCGGC-GGCGCTCTGGGGGCGGAGACAGCAAGCGGCTACGCGCGCGCAA 7643
| | | | |
QY 724 - - - - -AspValLeu 726
| | | | |
Db 7644 CGCTTCTCTGAGCGCTTTCGCGAGCATCGGCGGAGCGCTGGATTGACAGCGCTCGGT 7703
| | | | |
QY 727 GlyLeuGlyAspVal- - - - -Pro 732
| | | | |
Db 7704 GCGCTGGCGCGCTGGGCGGCGGCGATGGCCACCGATCAGGCGGCGAGCCACTCCA 7763
| | | | |
QY 733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet- - - - -TyrValI 748
| | | | |
Db 7764 ACAGCGCGCTGTCTCGGATGGCCCTCTCGCTTCTCGCTGGCGCGCTCGCGCTGCT 7823
| | | | |
QY 748 eAspValAspThr- - - - -GlyAspGluAspTyrPheGluThrTyrProSe 764
| | | | |
Db 7824 GGAGCAGCAGAGACACCGCTCACCGTCGCGGACATCGACTGGGCGCGCTTTCGCGCTTC 7883
| | | | |
QY 764 - - - - -GlnArgLysTyrPheGluCysAspPheValAspLysHisGlnArgTh 779
| | | | |
Db 7884 GTTCAGCGCGCTCGCGCGCGCTCTCTCGCGCAT- - - - -TTGCCGAGCGCGCGCGC 7940
| | | | |
QY 779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 799
| | | | |
Db 7941 TCTCGAGACCGAGCGCGCTCTCCGAGCATGCCCGGC- - - - - 7983
| | | | |
QY 799 uProValLeuGluAspArgAlaAspLysValLysMetArgGlnAlaIleGlyAspAr 819
| | | | |
Db 7984 -CCGACCTCTCTCGACAAGCTCCGAGCGCTTCGAGAGCGAGCAGCTT- - - - -CG 8033
| | | | |
QY 819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
| | | | |
Db 8034 TCTGCTCTCTGCTGCTGCGCACGAGCGCGCTCTCTCTCGCGCCACGAGGCGCGCTC 8093
| | | | |
QY 839 - - - - -ValGluPheArgGlnTyr- - - - - 844
| | | | |
Db 8094 CCATGTGACCCCGACAGGGCTTCTCGATCTCGTCTCTCGATTGCTCATGGCGCTCGA 8153
| | | | |
QY 844 - - - - - 844
| | | | |
Db 8154 GCTTCGCGCGCGCTTGCACAGCGCCACCGGATCAGCTCCGCGCACCTCGCGCTCGA 8213
| | | | |
QY 845 - - - - -ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
| | | | |
Db 8214 CCATCCCTCTCTCATCGAGTCCGCTCTTCTTGGCGACTGCG- - - - -TCGCCCCA 8264
| | | | |
QY 861 aLProPheLeuAlaGlyLeuProAspSer 870
| | | | |
Db 8265 CGCCCTCGCAGCAGGCTCTCCGTCGAGC 8293
| | | | |

RESULT 10
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5733759
; GENERAL INFORMATION:

APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patent In Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/457,335A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-335A-6

Alignment Scores:
Pred. No.: 0.000564 Length: 28958
Score: 178.50 Matches: 213
Percent Similarity: 29.8% Conservative: 89
Best Local Similarity: 21.0% Mismatches: 399
Query Match: 2.41% Indels: 312
DB: 1 Gaps: 42
US-09-913-878A-2 (1-1402) x US-08-457-335A-6 (1-28958)

QY 38 ProHisArgArgLys-----GluLeuAlaGluSer 47
DB 5663 CTTCCGAGGAGGAGATCCGCGACCTTCTCCAGCCCTACGAGCGCGAGCTAGCC---TCG 5719
QY 48 AspGluAspPheGlyArgHisAspLysileTyrArgAlaLeuAsnPheLeuTyrTipArg 67
DB 5720 CCGCCCTCAATGGGCTCTCTCCACCGTCGTCGTCGGCGATGAAGACGCGTGTGGAGA 5779
QY 68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaSer 87
DB 5780 TCGCCCGCCAGGCGGAGAGCTTCGACGAAACACACGCGCTCGCGCTCAGCCAGCGCT 5839

QY 88 Ser-----AsnTrpVal-----ProLysAla--- 94
DB 5840 TCCATTCCCGGCACATGGAGGATCTCGACGACTTCCGCGCGTCGCCAGAGCTCA 5899
QY 95 -----HisAlaAspProAspThr-----LeuProLysSer 104
DB 5900 CTTACCATCCGCGACGATCCCATCATCTCCAACGTCACCGCGCGCGCGCCAGCAGACC 5959
QY 105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
DB 5960 ACGAGTCGCTCGCGCGACTCTGGTCCGCCACGTTGCCACACCGTCCGCTTCTCTCG 6019
QY 125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
DB 6020 ACGCGCTAGTGCCTTC---ACGCGAAGGGGACGCTCTTCTTCGAGCTCGGCGCTC 6076
QY 145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
DB 6077 ACGCTGTCTCTCCGCCCTTGGCAAGACGCTCGACAGGACGAGGACGATGCGCCAT 6136
QY 165 GluProAla-----AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
DB 6137 GCGCCTTCTCTCCACCTCCGCAAGGAGCGAGCAGCAGCGAGGCGTTTACCGCGCGCG 6196
QY 182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
DB 6197 TCGCGCTCTCCACTCCG-----CAGGCATCAGCAGCGACT 6232
QY 202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
DB 6233 GGAGCGCTTCTTCGCCCTTCGCTCCAC----- 6262
QY 222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaProProValPro---Ile 240
DB 6263 -----GCAAGTCTCCCTCCCGCCTATGCTTCCAGC 6295
QY 241 AlaserAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
DB 6296 GCGAGCGCTTCTGCGCGAGCGCTCCAGCGCCGCG-----CCGACGTCA 6343
QY 261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
DB 6344 GCCACCTTGTCTCGCTCGAGGGGGGCTCTGCAAGACCATCGAGCGCGGGACCTCGATG 6403
QY 281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
DB 6404 CGCTAGCGGTGAGTCTCAGCTGAGCGGCGAGCGCGCGCGCGCTCGCCCTGCTCC 6463
QY 301 AlaProProSer-----GlnProArgLysLysArgProValAsp 313
DB 6464 TTTCCACCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGACGCTCGACGCTGCG 6523
QY 314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
DB 6524 GCTACCGTA-----TCACCTGGAAGCTCTTGACACCGCGCG 6559
QY 334 SerHisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThr 353
DB 6560 AAACACCCCGCGACCTCGCGCGCACCTGCTGCTGCTGCGCGCGCTCTGGACGAG 6619
QY 354 TyrTyrGluSerPheProSerSerGlyGlyGlyAlaIleProGluProSerArgSer 373
DB 6620 ---ACGCGTCTCCCTCCGCGCTCA-----CCGAGGCGCTCA 6652
QY 374 AsnGlyLeuAlaArgSerGluGlu-----SerAlaArg 384
DB 6653 CCGCGCGCGCGCGCGCTCGCTCGCTCGCTCGAGCCAGGCCACCTGGACCGGAGG 6712
QY 385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
DB 6713 CTC-----TCGCGGACATCTCGCCAGGCTTCGCGCGAGACCGGCCCGA 6757
QY 405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424

| | | | | | | | |
|----|------|--|------|----|------|---|------|
| QY | 539 | LeuIleProSerProThrSerThrSerProSerValPro-----ProValValSerLysGln | 557 | QY | 839 | -----ValGluPheArgGlnTrp----- | 844 |
| Db | 7181 | -----CCGGAGGACCAATTCATCACCAGCGGACCGGCCCATTTGGCGCTCACGTGC | 7234 | Db | 8094 | CAATGTCGACCCGACAAAGGCTTCTCGATCTCGTTCGATTCGCTCATGCGGTGCA | 8153 |
| QY | 558 | Pro-----GlyAlaValGluGluValIleGlnTrpLeuMetGlyGlnHisSerLeuVal | 576 | QY | 844 | ----- | 844 |
| Db | 7235 | CCCGATGGTCTCGTCAAGAGCGCTCAGCACCTCGTCTCATCA----- | 7279 | Db | 8154 | GCTTCGCCGCGGCTTGCACAGCCACCGGCATCAAGTCCCGGCCACCTCGCCTTCA | 8213 |
| QY | 577 | GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu | 596 | QY | 845 | -----ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV | 861 |
| Db | 7279 | ----- | 7279 | Db | 8214 | CAATCCTCTCTCATCGAGTGGCTCTTCTGGCGACTCGC-----TCGCCCA | 8264 |
| QY | 597 | PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPheAla | 616 | QY | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7280 | -----GCCGCGGAGCGCGGAGGCCCTG----- | 7303 | Db | 8265 | CGCCTCGGACGAGGCTCTCGGTGAGC | 8293 |
| QY | 617 | GluThrGlyLeThrPheArgProAspValPheLysThrArgSerValProAlaGlu | 636 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7304 | -----CGCGCTCGAGCTCCACGACGAGCTCTCGGCCCTCGCGCGC | 7345 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 637 | GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuGln | 656 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7346 | GCACACCTCCCGGTCGATGTCGCGACCGAATGCTGTGCCACGCTTCTTGAGC | 7405 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 657 | LeuAspAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeu | 674 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7406 | AGCTCGACCGCAAGGTCGAGGTCGCGCGCTGCTCCACGCGAGCGCATCGAACACC | 7465 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 675 | GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis----- | 688 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7466 | AGCTTCGTCGACGCCCTCTTCAGGATCTCGCGAGTTCTTCGCGCAAGTCG | 7525 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 689 | GlnIleArgHisLysThrAspLeuLeuSer-----ProSerGlyThrGlyGluVal | 706 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7526 | AAGTGCAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGAGCGCTTGTCTC | 7584 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 707 | MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg----- | 723 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7585 | TTTTCGTCGCGC-GCGCGTCTGGGCGCGGACAGCAAGCGCTACGGCGCGGAAA | 7643 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 724 | ----- | 726 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7644 | CGCCTTCCTCGAGCGCTTCCGAGCATCGGCGGAGCGTGGATTGACAGCGACGTCGT | 7703 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 727 | GlyLeuGlyAspVal----- | 732 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7704 | GGCTCGGCGCGTGGGCGGCGGCGATGCGCCACCGATCAGCGCGACCCACTCA | 7763 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 733 | SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----TrpValIle | 748 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7764 | ACAGCGGCTCTCGCGATGCGCCCTCGCTTCCCTGCGCGCTCGCGCTGCTCT | 7823 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 748 | eAspValAspThr-----GlyAspGluAspTrpIleGluThrTyrProse | 764 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7824 | GGAGCAGCAGACACCGCTCACCGTCCCGACATCGACTCGCGCGCTTTCGCCCTTC | 7883 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 764 | r-----GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh | 779 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7884 | GTTACGCGCGCTCGCGCGCGCGCTCTCGCGAT-----TTGCGGAGCGCGCGC | 7940 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 779 | rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnGlnLeuLe | 799 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7941 | TCTCGAGACAGCAAGCGCGCTCTTCGAGCATGCGCGCGC----- | 7983 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 799 | uProValLeuGluAspArgAlaAspLysValLysMetArgGlnAlaIleGlyAspAr | 819 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 7984 | -CCGACCTCTCGACAGCTCGGAGCGCTCGGAGCGGAGCGAGCTT-----CG | 8033 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| QY | 819 | gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro | 838 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |
| Db | 8034 | TCTGCTGCTCTGTCGCGCCACGAGCGCGCTCTCGCTCGGCCACGAGCGGCTC | 8093 | Db | 861 | alProPheLeuAlaGlyLeuProAspSer | 870 |

Alignment Scores: 0.000564 Length: 28958
Pred. No.:

RESULT 12

US-09-028-934-6
; Sequence 6, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/729,214
; APPLICATION NUMBER: 09-OCT-1996
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1506/CIP7
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-028-934-6

Score: 178.50 Matches: 213
Percent Similarity: 29.87% Conservativity: 89
Best Local Similarity: 21.07% Mismatches: 399
Query Match: 2.41% Indels: 312
DB: 3 Gaps: 42

US-09-913-878A-2 (1-1402) x US-09-028-934-6 (1-28958)

| | | | | |
|----|------|--|----------------------|------|
| QY | 38 | ProHisArgArgLys----- | -----GluLeuAlaGluSer | 47 |
| DB | 5663 | CCTCGAGGAGGAGTCCGCGACCTTCTCCAGCCCTACGAGGCGGAGCTAGCTAGCTCG | -----TCG | 5719 |
| QY | 48 | AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg | 67 | |
| DB | 5720 | CGCCCTCAATGGGGCTCTCTCCACCTGCTGCTGGCCATGAAGCGGCTGTGAGA | 5779 | |
| QY | 68 | LysAspAspSerLeuAsnGlnAlaGluAlaAsnPheIleGluAlaLysAlaAlaSer | 87 | |
| DB | 5780 | TCGCCCGCGCGCGGAGCCCTCGACGAAAGACACACGCTCGCGCTCAGCCAGCCCT | 5839 | |
| QY | 88 | Ser-----AsnTrpVal----- | -----ProLysAla-- | 94 |
| DB | 5840 | TCCATCCCGCGACATGACGAGATGCTCGAGACTTCCGCGCGCTCGCCAGAGCCCTCA | 5899 | |
| QY | 95 | -----HisAlaAspProAspThr----- | -----LeuProTrpSer | 104 |
| DB | 5900 | CCTACCATCCCGCACGATCCCATCATCTCCAAAGCTCACGCGCGCGGCCACGAGCC | 5959 | |
| QY | 105 | LysGluProProArgAlaAlaThrAlaGlyGlnTrpAlaLeuGlnThrValLeuLeu | 124 | |
| DB | 5960 | ACGAGCTCGCTCGCGCGACTACTGGGTCCGCCAGCTTCGCCACACCCGCTCGCTCG | 6019 | |
| QY | 125 | GluValLeuAsnArgPheMetProProAsnAsnThrProGlyArgThrPheGlyArg | 144 | |
| DB | 6020 | ACGGCGTACGTGCGCTTC--ACGCGAAGGGGACGCTGCTTCTCGAGCTCGGGCTC | 6076 | |
| QY | 145 | ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp | 164 | |
| DB | 6077 | ACGTGTCTCTCGCCCTTCGCAAGACGCGCTCGAGGAGGAGGAGGAGGAGGAGGAG | 6136 | |
| QY | 165 | GluProAla-----AsnValThrPheAlaAspProProLysArgSerLeuThrArg | 181 | |
| DB | 6137 | CGCCTTCTTCCACCTTCGCGAGGAGCGGAGGAGCGGCGGCTTACCGCGCGC | 6196 | |
| QY | 182 | SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal | 201 | |
| DB | 6197 | TCGGGCTCTCCACTCCG-----CAGGCATCACACCCGACT | 6232 | |
| QY | 202 | AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg | 221 | |
| DB | 6233 | GGAGCGCTTCTTTCGCCCTTCTGCTCCAC | 6262 | |
| QY | 222 | AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaProProValPro--Ile | 240 | |
| DB | 6263 | -----GCAAGGTCTCCTCCCGACCTATGCTTCCAGC | 6295 | |
| QY | 241 | AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla | 260 | |
| DB | 6296 | CGAGCGCTTCTGGCGCGACCCCTCCAGGACCCCGCGC-----CCGACGTCA | 6343 | |
| QY | 261 | ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr | 280 | |
| DB | 6344 | GCCACTTGTCTCGCTCGAGGGGGCTCTGGCAAGCATCGAGCGGGGACCTCGATG | 6403 | |
| QY | 281 | GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu | 300 | |
| DB | 6404 | CGCTCAGCGGTCACTCCAGCTGACGCGACGAGCGCGCGCGCTCGCCCTGTCTCC | 6463 | |
| QY | 301 | AlaProProSer-----GlnProArgGluLysArgProValAsp | 313 | |
| DB | 6464 | TTCCACCTCTCGAGCTTTCGCGAGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG | 6523 | |
| QY | 314 | AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys | 333 | |

| | | | | |
|----|------|--|--------------------------------|------|
| DB | 6524 | GCTACCGTA----- | -----TCACCTGGAGGCTCTGACACCGCGC | 6559 |
| QY | 334 | SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyThrSerPheSerThr | 353 | |
| DB | 6560 | AAACACCCCGACCTCGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 6619 | |
| QY | 354 | TyrTyrGluSerPheProSerSerGlyGlyGlyAlaIleProGluProSerArgSer | 373 | |
| DB | 6620 | ---ACGCGCTCCCTCCGCGCTCA----- | -----CCGAGGCGCTCA | 6652 |
| QY | 374 | AsnGlyLeuAlaArgSerGluGlu----- | -----SerAlaArg | 384 |
| DB | 6653 | CCCGCGCGCGCGCGCTCTCGCTTGGCTGAGCCAGCCACCTGGACCGGAGG | 6712 | |
| QY | 385 | SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys | 404 | |
| DB | 6713 | CTC-----TCGCGAGCATCTCGCGAGCTTTCGCTCAAGCCCTCGCGACCTCG | 6757 | |
| QY | 405 | PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe | 424 | |
| DB | 6758 | TTGCGGCGTCTCTCGCTTCTCGCTCGCGAGGCG-----CCCTCGAGACCGTC | 6811 | |
| QY | 425 | MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp | 444 | |
| DB | 6812 | CTGCCCTGCGCGGACTCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 6871 | |
| QY | 445 | SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg | 464 | |
| DB | 6872 | ACCTGAGCGCGCTTGTGCTTCTTCACGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCT | 6931 | |
| QY | 465 | GlyLysProPheProGluLysPro----- | ----- | 472 |
| DB | 6932 | CCCTGCGCCATCCCGCCAGGCGCATGCTGGGCTTGGCGCGCTCATCGGCTCGAGC | 6991 | |
| QY | 473 | -----ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer----- | 487 | |
| DB | 6992 | ACCCGACCGTGGGAGGTCTCGTACGCTCGCTGGGTTCGAGGAGCGCGCTGG | 7051 | |
| QY | 488 | -----LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn | 501 | |
| DB | 7052 | GCGCTTGTGCGGCCCTCGCGAGCGCCAGCAGACAGACAGCTGCTCTCCGCCCG | 7111 | |
| QY | 502 | ProAspAsn-----SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu | 518 | |
| DB | 7112 | CGGACTCTACCTCGCGCATCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG | 7165 | |
| QY | 519 | MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPheGluIle | 538 | |
| DB | 7166 | -----CGCGGACTTCACGC----- | 7180 | |
| QY | 539 | LeuIleProSerProThrSerThrSerProSerValPro--ProValValSerLysGln | 557 | |
| DB | 7181 | -----CCGAGGCGACCATTTCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 7234 | |
| QY | 558 | Pro--GlyAlaValGluGluValIleGlnTrpLeuMetGlyGlnHisSerLeuVal | 576 | |
| DB | 7235 | CCGATGGCTCTCGAGAGGCGCTCAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 7279 | |
| QY | 577 | GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu | 596 | |
| DB | 7279 | ----- | 7279 | |
| QY | 597 | PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla | 616 | |
| DB | 7280 | -----GCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 7303 | |
| QY | 617 | GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu | 636 | |
| DB | 7304 | -----CGCGCTCGGAGCTCCACGACGAGCTCTCGCGCGCGCGCGCGCGCGCG | 7345 | |
| QY | 637 | GluProValGluGluArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln | 656 | |

7346 GCACACCTCGCGCGTGCATGTCGCCGACGGAGTGTGTCGCCACGCTTCTTGAGC 7405
QY 657 LeuAspAenThrTrpGlnProHisLeuLysLeuPheSerArgIle-----GlnLeu 674
DB 7406 AGCTCGACCGCAGAGGTCGACAGTCCGCGCGTGTTCACGCGAGCGGCATCGAACACC 7465
QY 675 GlyLeuSerLysThrTyraIleMetThrLeuGluProHis-----688
DB 7466 AGCTCGCTCGACGCCCTCTTTCAGGATCTCGCGAGTGTCTCCGGCAGGTCG 7525
QY 689 GlnIleArgHisHisLysThrAspLeuSer-----ProSerGlyThrGlyGluVal 706
DB 7526 AAGGTGCAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCTTTGTTCTC 7584
QY 707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg-----723
DB 7585 TTTTCGTCGGGCGC-GCGCGTCTGGGCGGCGGACAGCAGCGGCTACGGCGCGCAAA 7643
QY 724 -----AspValLeu 726
DB 7644 CGCTTCCTCGACGCGCTTTCGAGCATCGCGCAGCGCTGGATTGACGACGCTCGGT 7703
QY 727 GlyLeuGlyAspVal-----Pro 732
DB 7704 GGCCTGGGCGGTGGGCGGCGGCGCATGGCCACCGATCAGGCGGCGACGCCACCTCCA 7763
QY 733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----TrpValIle 748
DB 7764 ACAGCGCGCTGTGCGGATGGCCCTCGCTTCCCTGCGCGCGCTCGCGCTGCT 7823
QY 748 eAspValAspThr-----GlyAspGluAspTrpIleGluThrTrpProSe 764
DB 7824 GAGCAGCAGCAGACCGTCACGTCGCGCATCGCTGCGGCGGCTTTCGCGCTTC 7883
QY 764 r-----GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
DB 7884 GTTCAGCGCGCTCGCCCGCGCTCTCTCGCGAT---TTGCCGAGCGCGACGCGC 7940
QY 779 rleuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 799
DB 7941 TCTCAGACACGAGGAGCGGCTCTCCGAGCATGGCCCGCC-----7983
QY 799 uProValLeuGluAspArgAlaAspLysValLysMetArgGlnAlaIleGlyAspAr 819
DB 7984 -CGGACCTCTCGCAAGCTCGGAGCGCTCGGAGCGGAGCAGCTT-----CG 8033
QY 819 rleuLeuAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
DB 8034 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8093
QY 839 -----ValGluPheArgGlnTrp-----844
DB 8094 CCATGTCGACCCGACAGGGGTTCTCGATCTCGTCTCGATTCTGCTATGCGCGTGA 8153
QY 844 -----844
DB 8154 GCTTCGCGCGCTTGCAACAGCGCCACCGCATCAAGTCCCGGCCACCTCGCGCTTGA 8213
QY 845 -----ValTyrgLuserTyrsSerArgAlaThrArgValSerHisGlyArgV 861
DB 8214 CCATCCCTCTCTCATCGAGTCGCGCTTCTTTCGCGCAGCTGCG-----TCGCGCA 8264
QY 861 alProPheLeuAlaGlyLeuProAspSer 870
DB 8265 CGCCCTCGGACGAGGCTCTCGTCTCGAGC 8293

RESULT 13
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: ShyTjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT

; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANC
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-30
Alignment Scores:
Pred. No.: 0.00461 Length: 14770
Score: 174.00 Matches: 277
Percent Similarity: 30.75% Conservative: 208
Best Local Similarity: 17.56% Mismatches: 551
Query Match: 2.35% Indels: 541
DB: 4 Gaps: 69
US-09-913-878a-2 (1-1402) x US-09-220-132-30 (1-14770)
QY 30 CysValAlaAspThrThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGlu 49
DB 4480 TGCACATTAAATACATCTGCGACGACATATAAAGGACAGAGTCAGATCAAGATGAT 4539
QY 50 AspPheGlyArgHisAspLysIleTyArgAlaLeuAsnPheLeuTyTrpArgLysAsp 69
DB 4540 GAGATTGAGAAACAGATAGACGACAGCTTCGCGATCCCTTAGCTTTAGTAAAGCGCTAC 4599
QY 70 AspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSerSerAsn 89
DB 4600 AGCTACTTCTGACTGAGCGCTGGA-----ATGATTGACGAGGTACAGGACGACACAGA 4650
QY 90 TrpValProLysAlaHisAla-----96
DB 4651 TCCCTCCCGACCATCTACTACTACAGCCATCTTTTCTACAGACCATACAGCTCTGG 4710
QY 97 -----AspProAspThrLeuProTrp-----103
DB 4711 ACAACAGCTCCGATTACAGTGCCTGGGCGACGCAAGTCAGGCTTCACCTTCCTTATCAAGT 4770
QY 104 ---SerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrVal 122
DB 4771 TCTTCTCTTAATACGCGCATCAGCTTCTCGGTTAAATCAATATGCTCTGTTTCGACA--- 4827
QY 123 LeuLeuGluValLeuAsnArgPheMetProProPro---AsnAsnThrProGlyArgThr 141
DB 4828 -----CCTTCTCCAATCAAAATCCACATTAGCGCGTCA 4860
QY 142 PheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLys 161
DB 4861 ACTACATCTTCAGTTAAATCCATTAGTACGTCGCGGATCTTCCAAATT---AGATCTTACGG 4917
QY 162 ArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
DB 4918 ACAATGCTTCGCGGATAAAACTGTGTGTTCACAAATCTCCATACATATCAGGTTTCC 4977
QY 182 SerAlaThr-----GlyProProLysHisGlyAlaAla 192
DB 4978 TCTGTGTACCTCGCTAGAGCTCCAGAGCTACCGGAGCTACGCCCTTAAAGGGGTGGA 5037
QY 193 IleProLeuLysPhe-----ProAspProValAsnThrGlySerLys-----206
DB 5038 TCCATTCTACGTTTCTCTCGAACCTCTCCAGTACTACAGCAGGGTCTCTTTTGGAG 5097
QY 207 -----ArgProSerLeuGluSerGluAsnLeuAsnGlnCysThr 219
DB 5098 AGGTCTCATTAATCTACTATGACACCCCTGCTCCCGCAATCAACATTAAATATGATTTC 5157

6913 AAAGAA-----GAGACTCACATAACCAACCAACCAAGTGGTTTATCAATCTCTCCA 6963
QY 862 PropheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPhe----- 877
Db 6964 CCAGCGGTGAGGTGCATCTCAAGAAATGAAGAAACCATGTCTGAGTCCATGACATCATG 7023
QY 878 ---LeuMetAsnSerGlyPheAspProLysLysGln-----LysTyrLeu 891
Db 7024 AAGGCTTTTCAGTCCGGGGGATCTTCCAAAGAACTGGCAGGTCTGTTTGAACATAAG 7083
QY 892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThr----- 905
Db 7084 TCGGCGATGTTCTCCAGATGTTTCCAGAGTCTGCTGTAACCTCAGCCAGCATGCGAGAG 7143
QY 906 ---LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyr 921
Db 7144 AAGGACACCAATGAACCCAACTGGCGGTATAATA----- 7182
QY 922 MetIleAlaAspPheTrpGlyValLeuGluGluValHisValGlyPheSerSer 941
Db 7183 -----GAAGTCCACATC----- 7194
QY 942 LysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 961
Db 7195 -----GAAAGAGTAACAGCTGAGCCACTGGAAGTCAATATTATAGA 7236
QY 962 ArgSerProAlaHisPheProSerAspIleGlnArgValA-gAlaValPheLysProGlu 981
Db 7237 GAAACCAAAAGCAT-----CCAGAA 7257
QY 982 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 1001
Db 7258 -----AAGAAATGATGATATATCAG-----AAA 7281
QY 1002 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1021
Db 7282 GACTTATCCGGGGAGATATT----- 7302
QY 1022 ValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLys 1041
Db 7303 -----AACCTAAAGATTTCTGCCA 7323
QY 1042 LysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys 1061
Db 7324 GAAAAACACGATGCTTTCTCT-----TGTTCCAGAGAACAGGTCAGCAGCAAGAA 7374
QY 1062 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 1081
Db 7375 GAAGAACTTACTGCTGAA----- 7392
QY 1082 LeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsn 1101
Db 7393 -----GAGTCA 7398
QY 1102 LysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly 1121
Db 7399 TTGCTTCTTATCTGGAGTCTTCCAGAGTAACACTCTCTGTGTCCTCCCAAGAAAGAT--- 7455
QY 1122 IleValPheAsnGluAlaSerTrpAlaGlnLeuArgGluLeuLeuGlyAlaLeu 1141
Db 7456 -----ACCGCCCTAGTCTGCTCACTC----- 7479
QY 1142 SerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHis 1161
Db 7480 ---ATATCTGATGACTCTTATAAAACATTGAAGCTTTTGAGTCAA-----CAC 7524
QY 1162 IleIleAspTyr-----LeuLysPheSer 1169
Db 7525 TCATAGATACCATGACGATGAGTGTGTCAGAACTAAGAGGGAGCTTACAGGTTGCT 7584
QY 1170 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAlaMetLysAlaLa 1189

7585 ---GAGAAATGCTTCTGTCCAGAAAGTAGATGTGTCTCATTTCT- 7626
QY 1190 LysAspThrGluAspGlyAla-----HisPheTrpAspProAspLeuAlaSerTyrTyr 1207
Db 7627 ---GATACTGAGATCGGTTACAGACCATGCGAGACCCCTAGC----- 7668
QY 1208 ThrPhePheLysGluIle-----SerAspLysSerArgSerSerAlaLeuLeuPheThr 1225
Db 7669 ---TCAGAGTTACAGGGGTCTGATAG---CGTCCAGAGAAATAATAGCCACT 7716
QY 1226 ThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLys 1245
Db 7717 GCCCCCAAAAAGAAATT-----CTCTCCAAATCTATAAGATGTTTCTGAAATGCT 7770
QY 1246 GluMetArgAspSerLysAspProTyrProValArgValAsnGlnVal----- 1261
Db 7771 GTAGGTAAAGTGTCTAAAGATGACATTTTGAATGACAGATGTTGCACTATTCTGCG 7830
QY 1262 ---TyrGluLysTrpCysAlaIleThrProGluAlaMetAspLys 1275
Db 7831 AATGTTAGTAGTCCAAACATCCATGTGGATGCGCTTTACTGAGACAGATTAGACAGA 7890
QY 1276 SerGlyAlaAsn-----TyrAspSerLysValIleArg----- 1286
Db 7891 GGTAGAGAGAAAGTTGATATATGAAGATAGGTGGACAGGACTGTGAAGGAGGCTGAAGAA 7950
QY 1287 ---LeuLeuGluLeuSer---PheLeuAlaAspArg-----GluMet 1298
Db 7951 AAAGTGAAGTGAAGTGTACAGTTTTTCTGCAAAACTGAAAGCTAAATGATGACTG 8010
QY 1299 AsnThrTrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerPro 1318
Db 8011 CAGTCCCCAGAGAAAGGACGCGCCCTAAATAATGCAAAAGATATTCTTCTCAAAGCCT 8070
QY 1319 Lys-----PheValTrpGlnMetAlaGlyArgGlnLeuAlaTyr 1331
Db 8071 ACCAGTAGAGCCCTGAGAAAGTGTACTGACAGAACTCTGCGCATCCATGATGAGTGG 8130
QY 1332 IleLysAlaGlnMetThrSerArgProGlyGluGly-----Ala 1344
Db 8131 GTTAGGCAAGACAGCATGCGCTGATGCAAGGCTTCCCAAGCCGAGGAGAGGCA 8190
QY 1345 ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLys 1364
Db 8191 CCAGTGTGCCAGCAGC-----CCAGAGAGATGGTTTCTCTCTCC 8229
QY 1365 GlnTyrVal-----AlaArgLeuGluGlyAspGlySer 1375
Db 8230 CAACAGACTGAGCAGCAGCAAGTCCACAGTGAAGCCAAAGGAAGT 8274

RESULT 14
US-09-426-436-1/c
; Sequence 1, Application US/09426436
; Patent No. 6225066
; GENERAL INFORMATION:
; APPLICANT: William R. Jacobs, Jr.
; APPLICANT: Barry R. Bloom
; APPLICANT: Graham F. Hatfull
; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/426,436
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/705,557
 FILING DATE:
 APPLICATION NUMBER: US/08/057,531
 FILING DATE:
 APPLICATION NUMBER: 07/833,431
 FILING DATE: February 7, 1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Pasqualini, Patricia A.
 REGISTRATION NUMBER: 34,894
 REFERENCE/DOCKET NUMBER: 96700/238
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-5995
 TELEFAX: (212) 286-0854 or 286-0082
 TELEX: TWX 710-581-4766
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 52297
 TYPE: nucleotide
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE:
 DESCRIPTION: phage genome sequence
 HYPOTHETICAL: no
 ANTI-SENSE: no
 FRAGMENT TYPE: not applicable.
 ORIGINAL SOURCE:
 ORGANISM: mycobacteriophage L5
 STRAIN: not applicable
 INDIVIDUAL ISOLATE: L5
 DEVELOPMENTAL STAGE: not applicable
 HAPLOTYPE: not applicable
 TISSUE TYPE: not applicable
 CELL TYPE: not applicable
 CELL LINE: not applicable
 ORGANELLES: not applicable
 IMMEDIATE SOURCE: mycobacteriophage L5 particles
 POSITION IN GENOME: entire genome
 FEATURE:

NAME/KEY:
 LOCATION:
 IDENTIFICATION METHOD:
 OTHER INFORMATION:
 PUBLICATION INFORMATION:
 AUTHORS: Hatfull and Sarkis
 TITLE: DNA Sequence, Structure and Gene
 TITLE: Expression of Mycobacteriophage L5:
 TITLE: A Phage System for Mycobacterial
 TITLE: Genetics
 JOURNAL: Molecular Microbiology
 VOLUME: 7
 PAGES: 395-405
 DATE: 1993
 US-09-426-436-1

Alignment Scores:
 Pred. No.: 0.00719 Length: 52297
 Score: 171.00 Matches: 388
 Percent Similarity: 32.70% Conservative: 192
 Best Local Similarity: 20.14% Mismatches: 524
 Query Match: 2.31% Indels: 513
 DB: 3 Gaps: 76

US-09-913-878a-2 (1-1402) x US-09-426-436-1 (1-52297)

QY 55 -----AspLysIleTyrArgAlaLeuAsnPhe 63
 Db 44606 ATGCGAGTGGGTAGCTTCGCGACGATGTCGGAAGACAAAGAGCGCTATTGTCGTGG 44547
 QY 64 LeuTyrTrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGlu--- 82
 Db 44546 TGAAGATGCTCCGGGACGACACTTA---GTGCTCGAGTTCGACCCATCCATCGAGCCC 44490
 QY 83 -----AlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAsp 99
 Db 44489 TAGGAAGGTATGCGCGCGCGGTTCAGGTACGTCCTCCCGGAGCATCGAGATGACGAT 44430
 QY 100 ThrLeu-----ProTrpSerLys-GluProProArgAlaAlaAlaThAlaGlyG1 115
 Db 44429 CTCCTGATCCGGGTGAACGAGCACACCACTACCGCGCGAGGTGAACCTCTCTGTCG 44370
 QY 115 nGlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProProAs 135
 Db 44369 T-----G3 44367
 QY 135 nAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgPr 155
 Db 44366 CCAGACGATCGAGGAGCTTCTCGGAGCGCTAAAGTACAGCCGCTTGTGTGAGCC 44307
 QY 155 oThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThr----- 170
 Db 44306 ACCCAGGCC---GCGCAACGAAAGAAACCCCGTGGCC-CTCCTCGGCTCGCAAAGCT 44251
 QY 171 -PheAlaAspProProLysArgSerLeuThrArgSer----- 182
 Db 44250 GTTCGCAGACAA-CCTCGTCCACACCTCACCACAGTCAATCCCGCATCATCCAGGTGGC 44192
 QY 183 -----AlaThrGlyProProIleHisGlyAlaAlaIlePr 194
 Db 44191 TGTTCGCGCAGAGGTGTGAGCGGTGTGAGGCCA----- 44155
 QY 194 oLeuLysPheProAspProValAsnThrGlySerLysArgProSerLeuGluSerCluas 214
 Db 44154 -----ACTGGA---CGAAGACCCAGCTCAGTTCTCTCTCG 44123
 QY 214 nLeuAsnGlnCysThr---LysArgAlaLysGlyLysLeuSerAspAsnValAlaAla 233
 Db 44122 TCTACCGCAGTCTCTGTGTAAAGACACCTCAGCACTTTATCG----- 44080
 QY 233 aAlaAlaProProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAl 253
 Db 44079 -----GCCTGTGTGCAATTTGATTCTTTCGCGCAACAAAGTCGAGTAGCACGGG 44027
 QY 253 aAsnThr-----ArgAspProThrAlaThrGlyHisArgArgAlaAspGlnVa 269
 Db 44026 GAATCGCTTTGATCCGACCATGATCCATCGACAAAGCATTTGTCGCGTCCGAAGCGTCAGCC 43979
 QY 269 lAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSerAlaCysAr 289
 Db 43978 GGATGGTGTCTCAGGATGACTCAGGCTG-ACCAATTCCCAAGGATCCGATTGACATGTGC 43920
 QY 289 gHisAsnGlnSerThrThrGlnSer-----SerPh 299
 Db 43919 GAGGTGCAAAACTGCTCAGTGTGCGTGTGACGGGGCGCTGAACGAGACGATAGGCAAG 43860
 QY 299 eGluAlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAl 319
 Db 43859 GGAAGATGCGCGCTCGTGACACACCAAGACATTTGTCGCGTCCGAAGCGTCAGCC 43800
 QY 319 aGlyHisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAsn-- 338
 Db 43799 AGTTGAACCAATACACACGTT-GCCCGAGGCGGTACAAAGCTG-GCTCGCATCCACAGGT 43742
 QY 339 -----GlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTrG1 356
 Db 43741 ATGGCTCGTCTCGCGGTGGTTCGCGAGGCG---ACCGCGTTTTCACACCGCTCGCTGA 43685
 QY 356 uSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLe 376

QY 35 ThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHis 54
 Db 44666 ACGCGAAACCTCAGCATAAGTCGAGGCTTTTCAGAGCTCCCGGATCACGGCGAGTAC 44607

Db 43684 GGTCTACGAGAGGCTCTGCGCAGGCGC-----CGCGAGATGAGCT 43643
 QY 376 uAlaArgSerGluGluSerGlnValGlnValHisAlaProValValAlaAl 396
 Db 43642 CGAACGGGCACAGAGATCTTCGCGAGGAG-----TACGCCAAGGACATCGGAGC 43592
 QY 396 aArgLeuArgAsnIleThrProLysPheProLysTrpLeuHisGluAlaProLeuAlaVa 416
 Db 43591 A---CTGTGCGAGCAAAACCCGAACTTCGAGTGGTGTTCGTCGGGCCCTACACGG 43535
 QY 416 lAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspLysSer 436
 Db 43534 TGAGCGTGACATCGAGCGCGGTTC---CACCTTGAGCTGGAGCAGGTAGAGAAGTTCAT 43478
 QY 436 uGlyLeuLysThrAspProSer-----TrpSerThr---AlaAspVal---- 450
 Db 43477 CGCGTGGCGAAGCAAGAGCCAGCAGATCTGGACTACCTCGGCGAGGATCTGCTC 43418
 QY 451 -ThrAspIleTrpLysThrLeuThrArgLeuAspAlaPheArgGlyLysPheProGln 470
 Db 43417 TACGGAGCGGTGGAAGACACCAACTCAAGAGAGTGTGAACAGCCCAAGCCGCCATCGA 43358
 QY 470 uLysProAsnAspValPheValThrAlaMetThrGlyAsnPhelLysSerLysGly--- 489
 Db 43357 GCTTCGG-----TTCAACATCGAGCTGGAGCGCAT 43328
 QY 490 -----SerAlaValValLeuSerAla-----ValLeuAs 499
 Db 43327 CCGCGTCGAGGTTTCATCGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43268
 QY 499 pTyAsnProAspAsnSerProThrAlaProLeuThrLeuValLysLeuLysProLeuMe 519
 Db 43267 CTACAGACCGGCAACTCGCGGCTGACGACTTCCAGCTCGGC-----GT 43223
 QY 519 tPheGluGlnGlyCysArgLeuThrArgPheGlyProAspArgPheGluLeuLe 539
 Db 43222 GTACGCGCTCGCGTCCGATGACC-----TATGACGTGA 43187
 QY 539 uIleProSerProThrSerThrSerProSerValProProValValSerLysGlnProGln 559
 Db 43186 GGCTCCGAGAGCGGCGACTTCTCATGCGCGGAGAGAGGATCAAGCGCAAGCCAC 43127
 QY 559 yAlaValGluGluValIleGlnTrp-----LeuThrMetGlyGlnHisSerLe 575
 Db 43126 TGCCTCCCTACGACTGACCGATTTGACGCGGAGCGGATCAGGAGAGGTTCATGAGT 43067
 QY 575 uValGlyArgGlnTrpArgAlaPheAlaLysAspAlaGlyTyArgLysProLeuAr 595
 Db 43066 CGAGGCTCGG-----ATCCAGGCTGGTGACTTCGAGCCTCTG-- 43030
 QY 595 gGluPheGlnLeuArgAlaGluAspProLysProLysProLysLysPhePh 615
 Db 43029 -----CTGAACCT-----GACAAATGCGGCTTC-- 43006
 QY 615 eAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAl 635
 Db 43005 -----TGTCATGTGAATTACAGCTCCCTGTTTAAATG-AGGTCA-----CCTTG 42960
 QY 635 aGluGluProVal-----Gln 640
 Db 42959 ACAGCCACCGATACCTACCTCCCTCTAAGCAGGACGAGGAATACAAAGAAACGCTC 42900
 QY 640 uGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuGlnLeuAspAsnAs 660
 Db 42899 GAACTCGACCCCGAGCGGGTTACAGTACGTGGAATGGGCGGATTCCTGACAGCCG 42840
 QY 660 nThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThr 680
 Db 42839 AAG-TGG-----CACGAGGAGCTGTACGCC-----TCGAGAAGTT 42805
 QY 680 rAlaIleMetThrLeuGlu-----ProHisGlnIleArgHisH 693

Db 42804 TCCGTTCCCGACCCAGAGGCTGCAGCGAGTTTCCCAAGCGGAGCGCGCTGCTCATCC 42745
 QY 693 sLysThrAspLeuLeu-SerProSerGlyThrGlyGluValMetAsnAspGlyValGlyA 713
 Db 42744 CGACCGAGATCGTGTATCGATTCGAGAGC-----GTCCGAGATGGAACGGCA 42694
 QY 713 rgMetSer-----ArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
 Db 42693 GGAGTTCGTGCGCGAGAGGTATCGAGCGCAACATGAACCG----- 42653
 QY 730 spValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp- 749
 Db 42652 -----CCGGCCACTTCATCCGC-3-AGTACCAGGAAGAGTTCGTCGTGTGACC 42607
 QY 750 --ValAspAspThrGlyAspGluAsp-----TrpIleGluThrTyProSerGlnA 766
 Db 42606 AACTTGTACAGCCAGAGAGAGACATGACGAACCTGGATCCAAACCCCTCGCTGA 42547
 QY 766 rgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValA 786
 Db 42546 GG-----TCACCTATAGCCCGCAGCAGAGCAGCTCGCTCCTACGG----- 42506
 QY 786 lAsrGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgA 806
 Db 42505 -----ATGCACCGCGCTGGATAC-----AAGG 42484
 QY 806 laArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnA 826
 Db 42483 GTCCGGGACCATGAAGATCTCAAGCTGCGCGGCACCCCGCTGATGAACACATGAGC 42424
 QY 826 rGlnPheSerGluGlnLysHisAla-----LeuAsnArgProValGluPheArgGlnT 844
 Db 42423 GAGCTTGGAGCGCGAGACGACGCGGCTCGCGCGCTCGCGG----- 42380
 QY 844 rpValTyrgLysSerTyrr-----SerSerA 852
 Db 42379 -ATTACAGCGCTTCGATCGATCCGAAGAGGTGAAGTGAAGGCCAGCCTCTACGACCA 42322
 QY 852 rgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnG 872
 Db 42321 GATCGTGCATCTCTACCGGACGATTCCTGCTGGGATTCGAG-CGTGACGACCGG 42263
 QY 872 luGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLys----- 886
 Db 42262 CAGAGGAGCTGCACGCGCTGTGTATCGAGCTGTAGCGCCCGCGCTCTGAAGAAAGCA 42203
 QY 887 -----LysGlnLysTyLeuGluAspIle- 894
 Db 42202 TGAGCAACACCTGTTTCAGTCCGATCTACATCGGACACACGCGCTGATGAGATCC 42143
 QY 895 -----AlaTrpAspLeuG 899
 Db 42142 GCAACCTCGCGGATGATGTCGAAGAGCAGCAGCCACCTGCGCAAGCGCTGGGACTCGG 42083
 QY 899 lnLysArgLysCysAspThrLeu-----LysS 908
 Db 42082 TCGTCGCAAGGACGACACGCTGTGGATCTCTAGGTGACATCTCTCTGGCTCCACCAAGG 42023
 QY 908 erLysLeuAsnIle-----ArgValGlyArgSerAlaTyrlleTyrr- 921
 Db 42022 GCGATATACCTCCCTCGGTTGATCTCGGATCGGCGCGGCGCAACCGCTGATCTCTGG 41963
 QY 922 -----MetIleAlaAspPheT 927
 Db 41962 GCAACACGATGGGCGCCACCGATGAACCGGATGCTCAACAGCTGTTGTTGCTGCTACT 41903
 QY 927 rpGlyValLeuGluGluAsnGluValHisValGlyPheSerLysPheArg----- 944
 Db 41902 GGATGTTGTTGAG-----CATGTGTCGAGCGAGCTGATCCGCGGTGCGCG 41855
 QY 944 ----- 944
 Db 41854 TGTAACGGGAGCGCTGGCGGCGACACAGATGTTCTGTAAGCCACTTCCCACTACGTCGGGG 41795

Qy 945 -----AspGluGluSerPheThrLeuLeuSerAspCysAspValLeuV 960
 Db 41794 ACCACACGAGGAGGACACACACCGAGTGGAGGCTCGGTACGATGCGAAGATCTGTA 41735
 Qy 960 aAlaArgSerProAlaHisPheProSerAspIleGluArgValArgAlaValPheLysP 980
 Db 41734 TC-----CACGGGACACGCGATTCCCGATGATCTCTCCGACACATCCACCGGA 41684
 Qy 980 roGluLeuHis-----SerLeuLysA 987
 Db 41683 GACAGATCCACGTGGGGATCGACGCTGGGGCGGCTGGTCTCTCGACGAGATCTATG 41624
 Qy 987 spValIle-IlePheSerThrLysGlyAspVal-----ProLeuAlaLys 1001
 Db 41623 ACCTGGTCAATCATATCCATCCACGAGGAGGAGGTGTACACACCTAGACAATCGCTTACATC 41564
 Qy 1002 LysLeuSerGlyLysP----- 1007
 Db 41563 CGAGGTGCGCGGGGATCCGCTACCCCGGCTGGCGGGCTTGACACGAGAAGGTACA 41504
 Qy 1008 -----TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle---ValAspGly 1024
 Db 41503 CATCTCCGACGAGGCGAGTATCTCTGCTCGCGCGCGCGGACCGGGAAGTCCGCG 41444
 Qy 1025 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 1044
 Db 41443 TTGCTCTCGCTACGCGCTCAAGTCGAGGTGCGGACGTTCTATTCTCGCGGACTCT 41384
 Qy 1045 ThrThrPheLysGluLeuMetAlaSerHisGlyThrGlySer----- 1058
 Db 41383 GATCGTTCCACGAGTCTCCGTTTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41324
 Qy 1059 -----AlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer----- 1072
 Db 41323 GCTACGCGGCTGTTCGCGAGGACACATCGAGGAGCGGTAGTGTACTACTAGACGAG 41264
 Qy 1073 -----PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 1088
 Db 41263 ATCCGATCCGTTCAACGATCAAGGATCAACGCTGCTG-----GACGAGATC 41216
 Qy 1089 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSer 1108
 Db 41215 GAGACGCTGTCCGCGATGATCGCGGTGACGCGGTGACGAGGATCTCCGAGCTGTATGCTGTG 41156
 Qy 1109 SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer 1128
 Db 41155 GACACATCACCAACGTCGCGACCGATTCAAGGAGGA----- 41117
 Qy 1129 TrpAlaGlnLeuArgGluLeuLeuGlyAlaLeuSerLeuProAspProMetTyr 1148
 Db 41116 -----GACACCGCTTCTCGGTCTCGAGTCTCGGTCTCGGTCTCGGTCTCGGTCTCG 41072
 Qy 1149 LysSerAspSerTrpLeuGlyArg----- 1156
 Db 41071 GAG-----ATGGCGGTGAGACAGGCTATGCTGCTGGGCTCCACACGTC 41024
 Qy 1157 -----GlyGluProThrHisIleIleAspTyrLeuLysPheSerIle 1170
 Db 41023 ACCGCGCGCACAAACGACGCGGACAGCGATCCCGTTGCTGGGATCAAGGTGAGATC 40964
 Qy 1171 AlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaLys 1190
 Db 40963 GGGCGT-----GTTCCTGAGTATGATCTCTCTCCACGA----- 40928
 Qy 1191 AspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrThrPhePhe 1210
 Db 40927 ---GTATCGACGGG-----TTGCGCGCGGATCTCAAGTCTCCACG---GTC 40883
 Qy 1211 LysGluIleSerAspLysSerArgSerAlaLeuPheThrThrLeuLys----- 1228
 Db 40882 AAGAACCGAGGGGAGTTCGACCGCTCGGACAGGATCTCGCTTCTTGGAGTTCGTC 40823

Qy 1229 -----AsnArgIleGly----- 1232
 Db 40822 GGAGACACTATGACATCAACGACTTCGGTCTCACTTAACCTTGACATCCACCAACAGAAAGT 40763
 Qy 1233 ---GluValGluLysGlyTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
 Db 40762 AGGACATGAGAGAATCAT-----CGCACCGGACTCATCGCCGATCGGC 40715
 Qy 1252 AspProTyrProValArgVal 1258
 Db 40714 GATCTGGGCTCTCCGCGTG 40694
 RESULT 15
 US-08-705-557-1/c
 ; Sequence 1, Application US/08705557
 ; Patent No. 630061
 ; GENERAL INFORMATION:
 ; APPLICANT: William R. Jacobs, Jr.
 ; APPLICANT: Barry R. Bloom
 ; APPLICANT: Graham F. Hatfull
 ; TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
 ; TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amster, Rothstein & Ebenstein
 ; STREET: 90 Park Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10016
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Word Processor (ASCII)
 ; CURRENT APPLICATION DATA: US/08/705,557
 ; APPLICATION NUMBER: US/08/057,531
 ; CLASSIFICATION:
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/833,431
 ; FILING DATE: February 7, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pasqualini, Patricia A.
 ; REGISTRATION NUMBER: 34,894
 ; REFERENCE/DOCKET NUMBER: 96700/238
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 697-5995
 ; TELEFAX: (212) 286-0854 or 286-0082
 ; TELEX: TWX 710-581-4786
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 52297
 ; TYPE: nucleotide
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE:
 ; DESCRIPTION: phage genome sequence
 ; HYPOTHETICAL: no
 ; ANTI-SENSE: no
 ; FRAGMENT TYPE: not applicable.
 ; ORIGINAL SOURCE:
 ; ORGANISM: mycobacteriophage L5
 ; STRAIN: not applicable
 ; INDIVIDUAL ISOLATE: L5
 ; DEVELOPMENTAL STAGE: not applicable
 ; HAPLOTYPE: not applicable
 ; TISSUE TYPE: not applicable
 ; CELL TYPE: not applicable
 ; CELL LINE: not applicable
 ; ORGANELLE: not applicable

Db 43186 GGCCTCGAAGACGGCGCACTACTTCATGCGCGGGAAGAAGGCATCAAGGCGAAGCCAC 43127
QY 559 yalavalgluValleGlnTrp-----LeuThrMetGlyGlnHisSerLe 575
Db 43126 TCGCGCTACGACCTCGCGGATGACCGGAGGCGGATCACCGAGAGGTTCATGAGGT 43067
QY 575 uValGlyArgGlnTrpArgAlaPheAlaLysAspAlaGlyTyrArgLysProLeuAr 595
Db 43066 CGAGGCTGG-----ATCCAGGCTGGTACTTCGAGCCTCTG-- 43030
QY 595 gclupheGlnLeuArgAlaGluAspProLysProLysLeuArgValHisPhePh 615
Db 43029 -----CCTGAACCT-----GACAAATCGGCTTC-- 43006
QY 615 ealagluThrGlyIleThrPheArgProAspValPheLysThrArgSerValAlProAl 635
Db 43005 -----TGTGATGTGAATTACAGCTGCCCTGTTTTTAAGTG-AGGTCA- 42960
QY 635 agluGluProVal----- 640
Db 42959 ACAGCCACGATACATCCTCCCTCTAAGCGACGACAGGAATAACAAGAAACGGTC 42900
QY 640 uGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAs 660
Db 42899 GAACTCGACCCCGACGCGGTTACACGTACGTGGAAATGGGCGGATTCCTTGACAAGCGC 42840
QY 660 nThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTy 680
Db 42839 AAG-TGG-----CACCGAGAGCTGTACGCC-----TCGAGAAGTT 42805
QY 680 xalatlMetLeuGlu-----ProHisGlnIleArgHisHi 693
Db 42804 TCCGTTCCCGACCCAGAGGCTGCACGCGAGGTTGCCCAAGCGGAGCGGTGCGTCATCC 42745
QY 693 slYsThrAspLeuLeu-SerProSerGlyThrGlyGluValMetAsnAspGlyValGlyA 713
Db 42744 CGGACGAGAGATCGTATCGAGTTCGAGGAGC-----GTCGAGATGGAACGGCA 42694
QY 713 xMetSer-----ArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
Db 42693 GGAGTTCGTGACGCGAGAGGTATCGAGCGACATGAACGG----- 42653
QY 730 spValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp- 749
Db 42652 -----CGGCGCACTTCATCCCGC-----AGTACCAGGAAGAGTTCGGTCTGTGACC 42607
QY 750 --ValAspAspThrGlyAspGluAsp-----TrpIleGluThrTyrProSerGlnA 766
Db 42606 AACTTGACAGCCACCGAGAGGAGACATGACGAACCTGGATCCAAACCCCGTCTGCTGA 42547
QY 766 rGlystrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValA 786
Db 42546 GG-----TCACCTATAGCCCGCACGACAGACCTCGCTCTACGG----- 42506
QY 786 laSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProValLeuGluAspArgA 806
Db 42505 -----ATGCAACCGCTGGATAC-----AAGG 42484
QY 806 laArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnA 826
Db 42483 GTCGGCGCAGCATGAAGATCTCAAGCTCGCGGCGCACCGGCTGATGACCAAGATGACG 42424
QY 826 rGlnPheSerGluGlnLysHisAla-----LeuAsnArgProValGluPheArgGlnT 844
Db 42423 GAGCCTTGGACGCGGAGACGACGCGGCTCGCGCGGTGCGGCG----- 42380
QY 844 rpValTyrGluSerTyr-----SerSerA 852
Db 42379 --ATTACGAGCGCTGTATCGATCCGAAGAGTGAAGTGAAGCGCCGCTCTAGACCA 42322
QY 852 rGlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnG 872

Db 42321 GATCTGTCGAATCCTCGACGGGACAGATTCCTGCTGGGATTCGAG-CGTGACGACGCG 42263
QY 872 luGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLys----- 896
Db 42262 CAGAGGAGCTGCACGCGCTGGTGTATCGAGCTGTATCGGCCCGCGCTTCTGAAAGAGACCA 42203
QY 887 -----LysGlnLysTyrLeuGlnAspIle- 894
Db 42202 TGAGCAACACCTGTTTCAGTCCGATCTACATCGACACAGCGCTGTATGGAGATCC 42143
QY 895 -----AlaTrpAspLeuG 899
Db 42142 GCAACCTGGCGGATGATGTGGAAGACGACGACCCCTGGCGAAGCGGTGGAGATCGG 42083
QY 899 lnLysArgLysCysAspThrLeu-----Lys 908
Db 42082 TCCTCGCGAAGGACGACGCGTCTGATCTCTAGTGCATCTCTCTGCTGCCTCCACCAAG 42023
QY 908 erLysLeuAsnIle-----ArgValGlyArgSerAlaTyrIleTyr- 921
Db 42022 GGCAGATACATGCCCTCGGCTGGATCTCGATCGGCGAGCGCAACCGCTGATCCTGG 41963
QY 922 -----MetIleAlaAspPheT 927
Db 41962 GCAACCAAGTGGGCCCCCACCAGATGAACCGGATGCTCACAGCTGGTGTGGCTACT 41903
QY 927 rpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArg----- 944
Db 41902 GGATGTTGTTCGAG-----CATGTGTCGACGCGAGCTCTGATCCGCGTCCGCG 41855
QY 944 ----- 944
Db 41854 TGTACGGGACGCTGGCGGCGCACACAGATGTTCTGTGTGAGCCACTTCCCATACGTGGGG 41795
QY 945 -----AspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuV 960
Db 41794 ACACACCACGAGGACCGACACCCAGTGGAGGCTCCGTGACGATGGCAAGATCCTGA 41735
QY 960 alalaArgSerProalaHisPheProSerAspIleGlnArgValArgAlaValPheLysP 980
Db 41734 TC-----CACGGGACACACGATTCGCCGATGATCTGTCTCCGACATCCACCGGA 41684
QY 980 roGluLeuHis-----SerLeuLysA 987
Db 41683 GACAGATCCACGTGGGATCGACGCTGGGCGCGGTGGTCTCTCGGACGAGATCTATG 41624
QY 987 spValIle-IlePheSerThrLysGlyAspVal-----ProLeuAlaLys 1001
Db 41623 ACCTGGTCAATCATCATCCACGAGGAGGAGGTGTACACACTAGACAATCGCTCTACATC 41564
QY 1002 LysLeuSerGlyGlyAsp----- 1007
Db 41563 CGAGGTCGCGCGGGGATCCGCTACCCCGCTGGGCGCGCTTACCAAGAGGTACA 41504
QY 1008 -----TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle---ValAspGly 1024
Db 41503 CATCTCCGACGAGCCAGCTAGTCTGCTGCGCGCGCCCGCACCGGAAAGTCGCG 41444
QY 1025 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 1044
Db 41443 TTCGTCTCGCTACGCGCTCAAGTCGAAGTGCCGACGTGTACTTCTCGCGCGACTCT 41384
QY 1045 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySer----- 1058
Db 41383 GATCGTTTCAACCGCTCTCCCGTTCGATCTCCATCTCAGCGGATGGTGTCTGGAACGA 41324
QY 1059 -----AlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer- 1072
Db 41323 GCTACGCGGGTGTTCGCGAGCAGAACATCGAGACGCGGTAGTGTACTCTACTAGACGAG 41264
QY 1073 -----PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 1088
Db 41263 ATCCCGATCCGGTTCATCTACAGGCATCACCCTGCTTG-----GACGAGATC 41216

Thu Apr 1 09:53:04 2004

us-09-913-878a-2.rni

QY 1089 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSer 1108
Db 41215 GAGACGGCTCTGGCCGCTATGACGGCTGTACAGGACTTCCACGCTTGATCGTCGTG 41156
QY 1109 SerLeuValGlyAsnLeuValAspGlnSerIysGlnGlyIleValPheAsnGluAlaSer 1128
Db 41155 GACAACATCACCAACGTCGCCGCCGATTCAACCGAGGA----- 41117
QY 1129 TtpAlaGlnLeuArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProwMetTyr 1148
Db 41116 -----GACACCCGTTCTCGGCTCTGGAGTCGCTGACTTACCTACAC 41072
QY 1149 LysSerAspSerTrpLeuGlyArg----- 1156
Db 41071 GAG-----ATGGCCGCTGAGACAGGCTCATCGTCTCGGGCTCCACACGTC 41024
QY 1157 -----GlyGluProThrHisIleAspTyrLeuIysPheSerIle 1170
Db 41023 ACCGGCCGCGACAAACACGGCGACAGCGATCCCGTTGTCGGGATCAAGGTCAGATC 40964
QY 1171 AlaArgProAlaIleAspIysGluLeuGluAlaPheHisAsrAlaMetIysAlaAlaIys 1190
Db 40963 GGGCGT-----GTTCTGAGATGATCTCACTCTCCACCGA----- 40928
QY 1191 AspThrGluAspGlyAlaHisPheTtpAspProAspLeuAlaSerTyrTyrThrPhePhe 1210
Db 40927 ---GTATCGGACCGG-----TTGGCCCGGACATGCTCAACGTCCTCCACG---GTC 40883
QY 1211 LysGluIleSerAspIysSerArgSerSerAlaLeuLeuPheThrThrLeuIys----- 1228
Db 40882 AAGAACCGGGGGGAAGTCGGACCGCTCGGGACAAGACTTCGCTTCTTGGAGTTCGTC 40823
QY 1229 -----AsnArgIleGly----- 1232
Db 40822 GGAGACACTATGAGATCAACGACTTCGGTCACTAATTGACATCCACCAACAGAAAGT 40763
QY 1233 ---GluValGluIysGluTyrGlyArgLeuValIysAsnIysGluMetArgAspSerIys 1251
Db 40762 AGCGAATGAGAGATCAT-----CGCACCGCACTCATCGCGGATCGGC 40715
QY 1252 AspProTyrProValArgVal 1258
Db 40714 GATCTGGGCTCTTCCCGGTG 40694

Search completed: March 31, 2004, 13:11:52
Job time : 675.141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:42:26 ; Search time 28 Seconds
(without alignments)
4816.447 Million cell updates/sec

Title: US-09-913-878A-2
Perfect score: 7397
Sequence: 1 MNPTPRKNSPVEEIIINRL.....YEVIGDDDFDGIQGTNGDYG 1402

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 328 | 4.4 | 1116 | T30828 | RNA-directed RNA p |
| 2 | 297 | 4.0 | 1114 | T30819 | RNA-directed RNA p |
| 3 | 296.5 | 4.0 | 1780 | T20695 | hypothetical prote |
| 4 | 279.5 | 3.8 | 1133 | T01920 | probable RNA-direc |
| 5 | 265.5 | 3.6 | 1215 | T11660 | probable RNA-direc |
| 6 | 232.5 | 3.1 | 3228 | T21381 | hypothetical prote |
| 7 | 191 | 2.6 | 5105 | T32650 | hypothetical prote |
| 8 | 182 | 2.5 | 1075 | T32253 | hypothetical prote |
| 9 | 177.5 | 2.4 | 966 | F84582 | hypothetical prote |
| 10 | 174 | 2.4 | 1226 | S15053 | hypothetical prote |
| 11 | 169.5 | 2.3 | 905 | G84582 | hypothetical prote |
| 12 | 163.5 | 2.2 | 4377 | A55575 | ankyrin 3, long sp |
| 13 | 160.5 | 2.2 | 1024 | T30868 | RhoA-binding prote |
| 14 | 160 | 2.2 | 1176 | JN0583 | myosin-light-chain |
| 15 | 159 | 2.1 | 529 | H84582 | hypothetical prote |
| 16 | 158.5 | 2.1 | 1736 | A47747 | tight junction pro |
| 17 | 153 | 2.1 | 1388 | T00063 | hypothetical prote |
| 18 | 153 | 2.1 | 1697 | T00079 | hypothetical prote |
| 19 | 148.5 | 2.0 | 1418 | S64918 | hypothetical prote |
| 20 | 148.5 | 2.0 | 1743 | T15893 | hypothetical prote |
| 21 | 144 | 1.9 | 2224 | 1 KFU05 | coagulation factor |
| 22 | 143.5 | 1.9 | 848 | S48273 | probable transcrip |
| 23 | 143.5 | 1.9 | 1316 | T20381 | KIAA0633 protein - |
| 24 | 143 | 1.9 | 967 | T20336 | hypothetical prote |
| 25 | 143 | 1.9 | 1285 | T14171 | ataxin-2 - mouse |
| 26 | 142.5 | 1.9 | 860 | JC4566 | chitinase (EC 3.2. |
| 27 | 142.5 | 1.9 | 3942 | T42730 | Bassoon protein - |
| 28 | 142 | 1.9 | 2409 | 1 A60979 | versican precursor |
| 29 | 141.5 | 1.9 | 1211 | T42330 | AP4 protein - mous |

30 141.5 1.9 3560 1 S02041 dystrophin, muscle
31 141 1.9 1576 2 S65774 homeotic protein H
32 140.5 1.9 1228 2 T18897 hypothetical prote
33 140.5 1.9 1516 2 P83085 conserved hypothet
34 140.5 1.9 3387 2 I40486 surfactin syntheta
35 140.5 1.9 3968 2 A44265 trithorax homolog
36 139.5 1.9 1207 2 T00378 KIAA0641 protein -
37 139.5 1.9 1367 1 S48478 glucan 1,4-alpha-g
38 139 1.9 2437 2 S53611 MIBP1 protein - ra
39 139 1.9 3685 1 A27605 dystrophin, muscle
40 138.5 1.9 890 2 I51618 nucleolar phosphop
41 138 1.9 2510 2 T42214 otogelin - mouse
42 137 1.9 653 2 D84582 hypothetical prote
43 137 1.9 2700 2 E84582 protein F21H11.2
44 136.5 1.8 1403 2 T11583 probable translati
45 136.5 1.8 1469 2 T19168 hypothetical prote

RESULT 1
T30828
RNA-directed RNA polymerase - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
C;Accession: T30828
R;Schiebel, W.; Pelissier, T.; Riedel, L.; Thalmair, S.; Schiebel, R.; Kempe, D.; Lott
submitted to the EMBL Data Library, October 1998
A;Description: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Ton
A;Reference number: Z20890
A;Accession: T30828
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1116 <SCH>
A;Cross-references: EMBL:AJ011576; PIDN:CAA09697.1
C;Genetics:
A;Note: RDRP
C;Superfamily: Arabidopsis probable RNA-directed RNA polymerase

| | | | | |
|-----------------------|--------|--|-----------------|----------------------|
| Query Match | 4.4%; | Score 328; | DB 2; | Length 1116; |
| Best Local Similarity | 19.9%; | Pred. No. 1e-11; | | |
| Matches | 211; | Conservative 149; | Mismatches 386; | Indels 314; Gaps 42; |
| Qy | 441 | DPSWSTARVDVDIWKTL-----YRLDAFRGKFPFPPKPPNDVVFV--TAMTCNFTS | 487 | |
| Db | 193 | DDQWVTTDTFTPSWIGLSSSLCIEFRNGVQLPNFSFFYKESMNQFIQTGFTFSFQ | 252 | |
| Qy | 488 | KGSVVLAVLDYNPNDSPTAPLYVVKLPMLFEQC--RLTRFGPDRFEILLISPTS | 545 | |
| Db | 253 | KLALVPIV-----HLPGIPLPKILFKISLTI--QHCFFPLALNF--NFFQLVDPRRN | 304 | |
| Qy | 546 | TSPSPVPVWSKQPGAVEV-----IQMLT-MGQHSVLGRQ----- | 579 | |
| Db | 305 | -----IACIEHALEKLYLKECCYDPVRWLTQYDVKGRQLPKSPPTLDDGLVY | 356 | |
| Qy | 580 | -----WRAPFAKDAVKKPLREFQLR--ABDPKPIK-----BRVHEF-----A | 616 | |
| Db | 357 | VRRVWVTPCKVYF---CGPEVNVSVRLRNYSEDINNFLRVSVFVDEEWKIHSTDLLPRA | 413 | |
| Qy | 617 | ETGITRPPDYFK-----TRSVVAEPVEQRTFVKYSQMLD--W-----LLQLDNNT | 661 | |
| Db | 414 | STGNGTDTDIYERILSTLRNGFIIGDKREFFLAFSSQLSDNVMWFASPPGLTANDIRT | 473 | |
| Qy | 662 | WQPHLKLFRI-----QLGSKTYAINTLEPHOIRHHKTDLL-----SPSGTGEVNDGV | 711 | |
| Db | 474 | WNGDFRQIRNVAKYAAFLGQSGFSRRETL---SVGRHEVEVIPDVACSLHGTNYIFSDGI | 530 | |
| Qy | 712 | GRMSRVAKIRDVULGLDVPSAVQGFSGAKGMWIDVDDTGDGDEWIEYTPSORKWECD | 771 | |
| Db | 531 | GKISADFAHRTVKCGLOQYTPSFQIRGYGKGVAVD-----PYSMK-----574 | | |
| Qy | 772 | FVDKHQRTLEVRSVASELKSAGLNQLL-----PVLEDRARDKVKMRQAIGRRLINDLQ | 825 | |

575 -----LSLRKSMKVESNNIKLDVLGWSKYQPCYLNRLQVLTLLSTLGVDDVLEQK 626
826 ROFSEKHALNPFVFRQWVYSYSRATRVSHGRVPLAGLPDSQEBTFLNMGSDP 885
627 NEAVQLDAILHDSLKAQEAELMS-----PGENTNILKEMLNCGYMP 669
886 KQKYLQDIADWLQKRCOTLKSXNIRVGRSAYITMIADFWGLVEEVEHVGSS----- 941
670 DAEPFLSMLOTFRASKLLDLTRTRIFIPNGRTWMGCLDSRTLEYGVQVQVSGARR 729
942 KRDEEESP--TLLSDC-----VLVARSAPHPFSDIQRVRAVFKPELHSLKDVIFST 993
730 QFESHPPNDSSANCDFILKGNVVAKNPCLPCHGDIRVLAVDVPALHVMDCVFPQ 789
994 KGDVPLAKLSGDDYDGMWVWVWPEIVDGVFNAEMPLEPDLRYLKQKXTTFKQLMAS 1053
790 KGRPHPNCSGSLDGDYFVWDPDLIPPRQVQSMYTPAPTQLDHD----- 839
1054 HGTGSAKQTYTDMIQKSFHALQPNFLGMCNTYKERLCYINNSVS----- 1100
840 -----VTIEVEEYF-----TNY-----IINDSLGIINAHVVPADRE 872
1101 -----NKPAIILSSLVNLDQSKOGIVFNEASWQRLRELGLGALSLPD-PMYKSDSWL 1154
873 PDWMSDPCQKLAQLFSIAVDPKTGVPAEIPS--QLRPKEYPDPMWKPDKPTVPSEKVI 930
1155 GRGEPTHTIDYLFKSTARPADAKELEAFHNAKAAKOTEDGAHFWDPDLASYTTFKELS 1214
931 GK-----LQKVKNK-----TPQASSIATFTRDVA 955
1215 DKSRSSALLFTTLKNRIGEV---EKEYGRLVNKNEMRDSKDPYVRVQVVEKWCATPPE 1271
956 RKSYSDDMIVDGFEDYIDFAFYKSEY-----DNKLGMLMDYGIKTE-----AEILSG 1004
1272 AMDKSGANYDSK-----VTRLELSFLADRENMTW-----ALLRASTAFKLYY 1314
1005 GIMKASKTDFRRKADBAIGAVVRCLR-----KEARAWFKRRSDIDDLAKASAWHYTV 1058
1315 HKS-----PKFVQWAGROLAYIKAOVTSRP 1340
1059 HHYTWGLYNEGLKRDHFIFPWCYVDQLIQIKAKARKP 1098
RESULT 2
T30819
RNA-directed RNA polymerase (EC 2.7.7.48) - tomato
C:Species: Lycopersicon esculentum (tomato)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30819
R:Schibbel, W.; Pellissier, T.; Riedel, L.; Thalmair, S.; Schibbel, R.; Kempe, D.; Lottst
Plant Cell 10, 2087-2102, 1998
A:Title: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato.
A:Reference number: Z20885; MUID:99055198; PMID:9836747
A:Accession: T30819
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1114 <SCH>
A:Cross-references: EMBL:Y10403; NID:G4038591; PIDN:CAA71421.1; PID:G4038592
C:Genetics:
A:Note: RdRP
C:Superfamily: Arabidopsis probable RNA-directed RNA polymerase
C:Keywords: nucleotidyltransferase
Query Match 4.0%; Score 297; DB 2; Length 1114;
Best Local Similarity 20.4%; Pred. No. 8.1e-10;
Matches 175; Conservative 133; Mismatches 323; Indels 228; Gaps 32;
562 BEVQIWTMQSHSLVGVGROWFAFFAKDAGYKPLREPOLRAEDPKPIKERVHFFAETGIT 621
386 EDIDNFRV---SFVDEEWKLYSTDL-----LPK-----ASTGSG 418
622 FRPDVFK-----TRSVVPAEPEVQRTFKVQSMLDWLLQLDNNTW-----QPHKLK--- 668

419 VRTNIYERILTLRKGFGVIGDKKFE--FLAFSSQL-----RDNVYMFASRPLGTANDI 471
669 -----FSRI-----OLGLSKTYAINTLPHQIRHKTDL---SPSGGEVMDGV 711
472 RAMWGDPSQLKNVAKYAARLGOSFGSSRETL---SVLRHEIEVDPVKVHGTSYVPSDGI 528
712 GRMSRVAKIRIDVLGLDVPNAVQORFSGAKGMVIVDDTGDDEDWIITYSQRKWEC 771
529 GKISGDFAHRAVASKCGLQVTPSAFQIRYGYGVGVGVPDPSMK---LSLRKSMKSYES 585
772 FVDKHQRTLEVRVAVSELKSAGLNLQLLPVLEDRARDKVMRQAIQDRILNDLQROFSEQ 831
586 NI-----KLDVLG-WKYYQPCYLNRLQITLLS-----TLGVKDEVLQKQKEAVDQ 630
832 KHALNPVPRQWVYYSYSRATRVSHGRVPLAGLPDSQEBTFLNMGSDPDKQKYL 891
631 LDAILHDSLKAQEAELMS-----PGENTNILKEMLNCGYKPADREPL 673
892 ODIADWLQKCKDTLKSXNIRVGRSAYITMIADFWGLVEEVEHVGFS---SKPRDRE 947
674 SMLQTRFASKLDDLTSTRIRIFIPNGRTWMGCLDSRTLEYGVQVQVFTGAGHGESDDL 733
948 ESF-----TLLSDCVDLVARSAPHPFSDIQRVRAVFKPELHSLKDVIIISTKGDVPL 999
734 HPFNNSRSTNSNFIKGNVVAKNPCLPCHGDIRVLKAVNVRAHLMVDCVWPPQKGRPH 793
1000 AKKLSGDDYDGMWVWVWPEIVDGVFNAEMPLEPDLRYLKQKXTTFKQLMASHTGSA 1059
794 PNECSGSLDGDYFVWDPDLIPPRQVQVMEYPPAPSIQLDHD----- 837
1060 AKEQTYTDMIQKSFHALQPNFLGMCNT---YKERLCYINNSVSKPAIILSSLVNLY 1115
838 -----VTIEVEEYFTNYIVNDSLGIINAHVVPADRE---EPDMMSDPCQKLAELFSIAV 890
1116 DQSKQGVFNEASWQRLRELGLGALSLPDPYKSDSWLGRGEPHTIIDLKFSIARPAI 1175
891 DFKTGVPAEIPS--QLRPK-----EYDFMDKPD-----KTSYISERVIGKLPKVK 936
1176 DKELEAFHNAKAAKOTEDGAHFWDPDLASYTTFKELSKSRSSALLFTTLKNRIGEV- 1234
937 DKA-----TPQASSIATFTRDVARRSYDADMEVDGFEYDIDEAF 974
1235 --EKEYGRLVNKNEMRDSKDPYVRVQVVEKWCATPPEAMDKSGANYD-----SKV 1284
975 DYKTEY-----DNKLGMLMDYGIKTE-----AEILSGGIMKASKTDFRRKDAEISVA 1023
1285 IRLELSFLADRENMTW-----ALLRASTAFKLYYHKS-----PKFV 1321
1024 VRLR-----KEARAWFKERNDDIDMLPKASAWHYTVHTPTVWGCYNQGLKRAHFISPP 1077
1322 WQWAGROLAYIKAOVTSRP 1340
1078 WCYVDQLIQIKKAKARNRP 1096
RESULT 3
T20695
hypotheical protein F10B5.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20695; T24512
R:Sims, M.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19310
A:Accession: T20695
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1780 <WIL>
A:Cross-references: EMBL:Z48334; PIDN:CAA8315.1; GSPDB:GN00020; CESP:F10B5.7
A:Experimental source: clone F10B5
R:Burton, J.
submitted to the EMBL Data Library, October 1995

A;Reference number: Z19901
A;Accession: T24512
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1780 <W12>
A;Cross-references: EMBL:Z66500; PIDN:CAA91312.1; GSPDB:GN00020; CESP:F10B5.7
A;Experimental source: clone T05C12
C;Genetics:
A;Gene: CESP:F10B5.7
A;Map position: 2
A;Introns: 23/2; 226/3; 305/2; 343/2; 462/3; 736/1; 922/2; 993/1; 1222/2; 1463/3; 1616/2

Query Match 4.0%; Score 296.5; DB 2; Length 1780;
Best Local Similarity 20.4%; Pred. No. 1.9e-09;
Matches 196; Conservative 129; Mismatches 348; Indels 289; Gaps 37;

QY 468 FPEKPNDFVFTAMTCNPES-----KGSAYLSAVLDVNPDSPT-----APLYL 512
DB 631 FDERQIDV-VTAFTTWQSRKIQYERLLKGESLQDVGLAKPLPKNCVSAKVIITPSRI 689
QY 513 VKLKLMEFQGRLLTRRFGDPFFILIPSPSTSPSPVVPVSKPGFAVEVIOWLTMGQ 572
DB 690 LLMAPEVMVN-RVVRREGPD----- 709
QY 573 HSLVGRQWRAFFAKDAGVKP-REFQLRAED-PKPIIKERVHFAETGITERPDVFK--- 628
DB 710 YAL-----RCVFRDDNLGLAIRDFSINNIDMSNIVTEGILTLKNGIQVADRVISFLG 764
QY 629 -----TRSVVPAEPPV-----EQRTEFRKVSQMLLQLLDNNTW 662
DB 765 WNSQMRDQCYLAPRVNALTGEVTGVEDIRVMGDFRDAISVPKMSRMGQCFTA- 823
QY 663 QHKLXLFRIQLGLSKTALMTLEBHQIRHKKHTDLLSPSGTGEVNN-----DGVGRMRS 717
DB 824 QP-----TVYSVKNHIVENIQVRLERHW--IVEPDIEGGVENKCYFGDGGGRISIK 875
QY 718 VAKRIKRVLDGVPVSAVQGRFSAGKMWVID--VDDTGDESDWIEETVPSQKWCDFVDK 775
DB 876 LATHISKILQKEVPACQVRPKFGKILVIDPTDIDLIINPKVIFRKSQKQFGEGGEL 935
QY 776 HORTLEVSASSELKAGLNQLLVLEDRADKVKRQ----- 814
DB 936 QDEYLVKRYAMP-SPVCLNRPFITL-----DQVSEKQASRRITNRVHYLRELC 989
QY 815 AIGRLINDLQRFSEQKHALNRVPEFRQWVYVSSRATRVSHGRVFLAGLPDSQBEET 874
DB 990 SLNKLINENQ-----AAELNRTNLAIWN--NAASKEA----- 1022
QY 875 LNFLMNSGDPKKQKYLQDIAWDLQKRCOTLKSKLNI-----RVGRSAIYMIADFWGLV 930
DB 1023 -----GFELSVDPILRDMLFSIYRYNIHHISKAKIFLPPSLGRS--MYGVVDETGLL 1073
QY 931 ENEVHVGFSSKFRDEBESFTLLSDCDVLVARSAPHRPSDIQVRVAVFKPBLHSLKDVII 990
DB 1074 QYGVFIQYSPSIR-OTSNRPILKTGKVLITKNPCHVPGDVVRVDVAVWQPALAHLVDVV 1132
QY 991 FSTKGDVFLAKLGGDYGDMAWVCWPEIVDGFVNAEMPLEPDLRYLKXKRTTTPKOL 1050
DB 1133 FPGHPRHPDPMAGSLDGDGEYSIIWQEMLLDY-NEEAMVFP----- 1175
QY 1051 KASHGTGAKEQTYTDMIQKSFHALQPNFLGCTNYKEELCYINNS-----VSNKPAIIL 1107
DB 1176 --SSAAAEDEKPTDDWVEFFLRY-LQDSIGRMNH--AHLAVADLHGLFHENCHIAL 1230
QY 1108 SSLVGNLVQSKQGIQVFNESAQAQRLRELLGGALLSPDPMYKSDSWLGRGEPTTHIDYK 1167
DB 1231 KCAVA--VDFPKSGVPAEPLS-----SFEQCEMTPDYMMSGGKPMYSTRLN 1275
QY 1168 FSIARPA--IKLEBAFNANMAKADTEDGAHFWDPDLLASYTTFKISKSRSSALLFT 1225
DB 1276 GQLHARKARKEVEVLEEF-----ETRGVSF--- 1299
QY 1226 TLKRIGEVEYGRLL-----VKNKEMRDSKDPYVRVNVQVKEKCAITPEA 1272

DB 1300 -----EREYDKLICPEDVDVFFGNEIKLVQTLTLDREYVDNRWQQLLDEY-GIEDEA 1349
QY 1273 MDKSGANYDSKVRILLELFLADRENMTWALLRPASTAFKLYHK-----SPKXVWQAGR 1327
DB 1350 SVVSG-----HAASIKRLAGMERDDYSFYHTDKVSELRYEKLYAVFRAKFFEBFGGE 1401
QY 1328 QL 1329
DB 1402 EI 1403

RESULT 4
T01920
Probable RNA-directed RNA polymerase (EC 2.7.7.48) - Arabidopsis thaliana
N;Alternate names: protein T22B4.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 10-Dec-1999
C;Accession: T01920; T08192
R;Strong, C.; Graves, T.; Duckels, G.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of A. thaliana P2P3.
A;Reference number: Z14455
A;Accession: T01920
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1133 <STR>
A;Cross-references: EMBL:AF080120; NID:G3600045; PID:G3600048
A;Experimental source: cultivar Columbia
R;Bevan, M.; Pohl, T.; Weissenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16098
A;Accession: T08192
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1133 <BEV>
A;Cross-references: EMBL:AL049876
A;Experimental source: cultivar Columbia; BAC clone T22B4
C;Genetics:
A;Map position: 4
A;Introns: 184/3; 820/2; 870/3
A;Note: F2P3.11; T22B4.110
C;Superfamily: Arabidopsis probable RNA-directed RNA polymerase
C;Keywords: nucleotidyltransferase

Query Match 3.8%; Score 279.5; DB 2; Length 1133;
Best Local Similarity 21.8%; Pred. No. 9.9e-09;
Matches 156; Conservative 103; Mismatches 245; Indels 211; Gaps 31;

QY 698 LSPSGTGEVMDGVGRMSRSVAKRIKRVLDGVPVSAVQGRFSAGKMWVIDV----- 751
DB 544 VTTDGADYCFSDGIGKISLAFKQVAKQKGLSHVPSAFQIRYGVYGVIAVDRSSPRKLS 603
QY 752 -----DTGDE-----DWIETVPSQKWCDFVDKH-----QRTLEVSASSELKSA 794
DB 604 LRDSMLKFDNNRMLNVTWTESMP-----C-FLNREIICLLSTLGIEDAMFEAMQA-V 655
QY 795 NLQLL-PVLEDRARDKVRQRAIGRLINDLQRFSEQKHALNRVPEFRQWVYVSSRA 853
DB 656 HLSMLGNMLEDR-----DAALNVLQK----- 676
QY 854 TRVSHGRVFLAGLPDSQBEETLNFMLNSGDFPKQKYLQDIAWDLQKRCOTLKSKLNR 913
DB 677 -----LSG--ENSKNLLVKMLLQGYAPSPSPYLSMLLRVHHESQLSELKRCRL 724
QY 914 VGRSAIYMIADFWGLNEVHVGFSS-----SKFR--DBEESFTLLSDCDVIV 960
DB 725 VPKGRILICGDMGMLLEYGVYVVRTLTAKELSRDQSYFRKIDETSIVI---GKVV 781
QY 961 ARSPAHPSPDIQVRVAVFKPELH-----SLKGVIIISTKGDVPLAKLGGDYGDMAWVC 1016
DB 782 TKNPCLHPGDIRVLDAIY--EVHFEKGYLDICIIIFPKGERHPNECSGGDLGDGQFFVS 839


```
QY 758 WJET-----YPSQKWECDFFVDKHQRTLEVRVSASELSKAGLNQLLPVLED 804
Db 2306 WALVNGVTSKMAKCLFRPSQIKFQAKAISGDQ--IEMVTFSSAVLVA-LNKPFINIL-- 2360
QY 805 RARDVKVQRQATG-----DLINDLQRFSEQKHALNRPVFPQWVYESYSS 851
Db 2361-----DQSENQSLDCHKRTSRIEELMDRQILISFAKOMNEETFCRNKLKEFFR----- 2409
QY 852 RATRVSHGVRVPLAGLPDSQEBTLNFLMNSGDPKQKYLQDIAMDLOKRKCDTLKSKLN 911
Db 2410---RIDIDNLRMTWGFLLSSSEFFPSLIKA-----SIKFSITKQLC---KQGIQ 2452
QY 912 I--RVGRSAYIYMIADFWGVLEENEVHVGFSKF-----RDEBESP--TLLSDCDVLVA 961
Db 2453 IPSELGRS--MLGVVDETRGLQYQGIQFYQYTKNYKKLPPRDSNNKVKHGSBIVTGTVLLT 2510
QY 962 RSPAHFPSPDIQVRVAVFPELHSLKXDVIIIFSKGDPVLAKKLGGDYDGDMAWCMWDEI 1021
Db 2511 KNPCIVPGDVRIFEAVDIPELHMCMDVVVPHQGRPHPDDEWAGSDLDGDSYVIMWQEL 2570
QY 1022 VDGFNVAEMPLEP-DLSRYLKXDKTTFKQLMASHOTGSAAKEQTYDMIQSFHFA-LQP 1079
Db 2571 L-----LERNEEFFDFA--VERIKVPY-----DREKLDVLMREFVYTVLKL 2609
QY 1080 NPLGCTNYKERLCVINNS---VSNKPAIILSLVGNLVDSKOGIVFNEASWAQLRRE 1135
Db 2610 DSVGOISN-----SLHNSDQGLNSRVCMDLAKNCOAVDFTKSG----- 2650
QY 1136 LGGALSIPDMYKSDSLWGRGEPTTHIIDLKFSIARPAIDKLEAFHNAMKAARDTEG 1195
Db 2651-----QPPDPL--ETKW--RADP-----VTFEVIPPENPERIPDEH-----MGNERS 2688
QY 1196 AHFWDPDLASYVTFPEKISDKSRSSALLFTTLKRIGEVEXEYGRUVKNKEMEDSKDYP 1255
Db 2689 PMYVSPLCG--KLRE-----FQIDN-----VIKISEERD--BOYN 2722
QY 1256 VRVNVQV-----YEKWCATPEAMDKSGANYDSKVIRLLE 1289
Db 2723 IELDETIFTGFERYM-----ESAQQLSSYNGQLRSIME 2757

RESULT 7
T32650
hypothetical protein F39C12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T32650
R:Chisoe, S.; Sansone, J.
Submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F39C12.
A:Reference number: Z21206
A:Accession: T32650
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5105 <CHI>
A:Cross-references: EMBL:AF039043; PIDN:AAB94194.1; GSPDB:GN000028; CESP:F39C12.1
A:Experimental source: strain Bristol N2; clone F39C12
C:Genetics:
A:Gene: CESP:F39C12.1
A:Map position: X
A:Introns: 19/2; 47/3; 103/2; 142/2; 174/2; 324/2; 362/3; 436/3; 494/1; 523/1; 923/3; 98
88/3; 491/3; 5020/3

Query Match
Best Local Similarity 18.4%; Score 191; DB 2; Length 5105;
Matches 303; Conservative 207; Mismatches 557; Indels 578; Gaps 81;

QY 85 AASNNWPKAHAD---PDTLPWSK--EPRAATAGQWALQTLVLELNR-----FMPPPN 135
Db 2841 AESEKVPVEKVDPSLTPPAASVPADPPKMPADQWRAAMWARRINGASKPVFVAPPS 2900
QY 136 NTPGRTFORTLGGPSGLSRPSTNTKXKDEBPANTVFADPPKSLTR----- 181
```

```
Db 2901-----ALSKTESLKEASNTIQESASTTRVQPALVVSHSTPEPVQLTRMAAAKSEKEYKT 2955
QY 182-----SATGPIHGAALPLKFPDPVNTGSKRSLSELENLNOCTK-----RAKGKLS 228
Db 2956 PVAKPKATAPP-----PITPKAASIRATQKLHSELTKSEASNVSSSGSKSQ 3005
QY 229 NVAATAAP-----PVP-IASALD-KVPTRRH---ANTRDPTATGHRRADQVDSFDT 274
Db 3006 QADVQSRPKLSQVQEKRSVPKVKPLDPKGPQKEQSKKNEKVVVPSNTVTLDDQASSSN 3065
QY 275 SQG-----TSYGSVFSACRHNOSTTOSSEAFPSQPR 307
Db 3066 QQPGGAKPSASIFALFORKEKKEKMTSTAESIPSPENQEQATBETNRSSILPR 3125
QY 308 EKEPVDATVEAGHLIE-SPS-----KGRTTKSHIDNOPLSSSSSQOETSFSTYES 357
Db 3126 IKXP-----ETQLIKRSPSTDVPKPCVVGAIVK-FVDGKPISNS----- 3165
QY 358 FPSGCGEGALPEPSRNGLA-----RSEASASQVQVHAPVVAARLNTWPKPKWLHEA 412
Db 3166 -PSYKPMARRPKKADSAADSGDLGSDGSGTS-----HEG 3200
QY 413 PLAVAMEVTRLFMHCKVDLEDESLGLKYPDSWSTARDVTDIWKTLYRLDAFRGKPPPEK 472
Db 3201-----C--LKDEENSQK-KPACTV-----RFGPSVEPEPEPAP 3231
QY 473 PNDVFTVMTG---NPESGKSAVV-----LSAVLDYNDPNSPTAPLYLKLPLMF 520
Db 3232 POPVEIPKEVPTPEISFESLAQMAQKTLRBEASKILEHYFVTEQTSPTAEPVSPT-- 3289
QY 521 EQCRLTRFRFGDPRP-----EILPSPSTSPSPVPPVVSQ 557
Db 3290 KSGSKLS-PFGLSPFGKVPKNTAQTSQTLVLESTSPPELLTPNPEISNTASTSSSQ 3348
QY 558 PGAVEE-----VIQWLTMGHSLVGRQWRAFFAKDAGYRKLPRF---QLRAEDPKP 606
Db 3349 PKSLSEKITATQMIQ--TRG-----RPIKLSHRRQLKIGDPS 3386
QY 607 IIKERVHFFAETGITFRPDVFKTRSVVPAEPEVQEOTEFEKVSQMLD-----WLLQL 657
Db 3387 L-----QLSGDSSLPR-----PAQSPKPPQLQEVFGVQTYNPPQRYKTYRL 3430
QY 658 DNNWTQPHLK---LFSRI-QLGLSKTVAIMTLEPHQIRHHKTDLLSPSGTGEVMN--- 708
Db 3431 -GREWKPPPKAKTPLSIRVPSQSVSTSTPAIDLSAKEV-----LLAPIETUKMKRKRKP 3483
QY 709-----DGVGRMSRSVAKIRIDVLGLGVPSAVQGRFGSAGKMWIVDDVDTGD--- 755
Db 3484 KPMETTLVLDGTEHPFKKLTAEMLQ--LLSRQPKSKRKNKKGTKTWSVIGCDENDDSTK 3541
QY 756 ---EDWIETPSQK-----WECDFVD---KHQ-----RTLE 781
Db 3542 MQHNKLLBEPDLKKSLLITRQLDKXMITARAVFNPRFTDSSSTKHELLRNKPKFETTRK 3601
QY 782 VRSVAS-----ELKSAGLNQL---LPVLEDRADKVQRQAIGDLRLNDLQRFSEQKHA 834
Db 3602 LKRLANNROOKTRMSGIHSQPKRDVAISAIROVQVFRFQSTGKGMNPIISGP-KPKSI 3660
QY 835 LNRPEFROWVYESYSSRATRV-SHGRVPFLAGLPDS-----QETLPLMNSGDPKPKQ 888
Db 3661 LKPPKEPATWLF-----RIDSEGNKLYMKLGPSSDWQKNDKCNPM--SAMDPRKK 3710
QY 889 K---YIQDIADWLQKRKCDTLKSKLNIRVGRAYIYMIADFVGVLBEENVHVGFSKFRD 945
Db 3711 KNSYLYK--RW-----AKWAKYVK----- 3728
QY 946 ESESFTLLSDCDVLVARSPAHFPPSDIQVRVAVFKPELHSLKDVIIIFSTKGDVPLAKKLSG 1005
Db 3729 --XNMTKLFETNQLTRILGAP-----KDSHP-KDVLVDPKADAPPSPR--- 3770
QY 1006 GDYDGDMAWCVDP-----EIVDGFVNAEMPLEDLSRYLKD-----KTFKOLMASH 1054
Db 3771 ----SLSWTTVSSNASLNSLLDA-LNTVGSDEPSTSGIVTDIAPPVIVKEQFRKAKSR 3824
```

```
QY 1055 GTGSAKQQTVDIMIQKGFHFALQPNFLGMCNTYKERLCY-----INNSVSNKPAI 1105
D 3825 PTQQTQSKK--YPTVYLGNTVARSPSSSTSLVNLDEREAVSLAQDSEIYEVYVQSEPV 3882
QY 1106 I-----LSSLVGNLVDSKQGVINEASNAQLRELLG-----CALSLPD--P 1146
D 3883 VETTKRRFDPFVRPISALFKAPTSSPEVGEADSRSPRTERTLRLGFRRSPTMPPEFST 3942
QY 1147 MYKSDS-----WLGGEPTHTII-----DYLKFSIARP-----AIDKELEAFNNAKAKDT 1192
D 3943 RHVSSSPTRRRANRGFVTSVQSGSAEDILALMQAOPDALENMBEQVVKSPPIQLQAPV 4002
QY 1193 EDGAHFMDPLA--SYTFFKEISDKSRSSALL-----FTTLKVRIGEVEKE 1237
D 4003 LNKGMWRPKLKNLSLTQFKELKEESKKLVLYLDQNKKNKSYSTLTITFLNLDQ 4062
QY 1238 YGRL-----VKNKEMRDSKDPYPVNVQVYKWCALITPEAMDKSGANYDSKVIRLLELS 1291
D 4063 YMRNLVPRPKLNTDLLGKYDPSLRVS-----SSRSGLQRSTRLI--VKKI 4108
QY 1292 FLADREMTW--ALLPASTAFKLYHKSPFVVMQAGROLAYIKQMTSRPGEAPALM-- 1348
D 4109 RCQOLEKWKRLIATSTNQQL-----TNGELVTKAKASKMKQASNRKAKKVLVL 4157
QY 1349 -----TAFWY-----AGLMPDKKFTKQYVARLEGD-----GSEYDPDPVVEV 1385
D 4158 YKFKTCTFKYRVQVITKSAKILADQKKQLINHEEKETPTIITIGDEDDENEIVDL 4217
QY 1386 L-----GDDPFDGI 1394
D 4218 LSGNEDSANQNADYDGDNDDAI 4242

RESULT 8
T38253
Hypothetical protein SPAC23C11.15 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38253
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21781
A:Accession: T38253
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1075 <BR>
A:Cross-references: EMBL:Z98559; PIDN:CAB1171.1; GSPDB:GN00666; SPDB:SPAC23C11.15
A:Experimental source: strain 972h-; cosmid c23C11
C:Genetics:
A:Gene: SPDB:SPAC23C11.15
A:Map position: 1
A:Introns: 11/1; 71/2

Query Match 2.5%; Score 182; DB 2; Length 1075;
Best Local Similarity 18.7%; Pred. No. 0.0086;
Matches 193; Conservative 157; Mismatches 359; Indels 322; Gaps 51;

QY 494 LSAVLDPNDSPTAPVL-VKLKPLMEQCRTRRRGDRPFPHILPS-PTSTSSVP 551
D 81 ISVILRDYDPLEYLNIFLPSYKLLNSGANFTLQF-----TTPSGVSTPTSYV 132
QY 552 PVSQKQGAVEREVIQWLTMGQSHLVGRQWRAFFAKDAGYRKPLREFQ----- 598
D 133 ATYNDLPCTYHRAIGFVSVRVALLSNPEQFKLQDS-----LKKFNSECSLSELOTIV 187
QY 599 ---LRADPKPIKERVHFETGTF--RPDV--FKTSVVPAAEPVEQRTFVKVQMLD 652
D 188 TSLLAHPS--LAHEFHNFLPSSIFPGKPLGSPFLRGI-----QSSQFTLSNISD 237
QY 653 WLQLDNNNTQSHLKLFSRIQLGLSKTYAINTLEPHQIRHKTLLSPSGTGEVMDGVG 712
D 238 LLSQ-----SRPNLSP--FSLHLSNE--- 256
```

```
QY 713 RMSRSVAKRIRDVL-----GLGVPSAVQGRFG---SAKGM--VID 749
D 257 --SSDFKVKVNLDTVETVHEFLKLLNLYVQGI--DRNILVSRGFGFLKSNGLWSPLS 314
QY 750 VDDTGDGDIETYPSSORKWECDFVQKHQRTLEVRVSASELKSAGLNQLQLLVLEDEARDK 809
D 315 L7SLSPPEFLSVY-----NSACDFPECGSPSYRLPVEERINISCS 354
QY 810 VKMRQAIGRLIND-----LQR-----QFSEQKHALNRPVEFRW 844
D 355 GRDDFAWG--ILNDDVWSHPTWASBESGFIQRTKYBEAMTKLEERYEYFDRHTEATSW 412
QY 845 VYESYSSRATRSHGRVPFLAGLPDSQETLNFLLNMSGDPKK-----QKYLQD 893
D 413 TIKSLKKIQNRINE-----LPBEERTYTLSEGLGLPSKSIYKTKIKLVYVTEHABE 464
QY 894 IAWDLQKRKCDTLK-----SKLINR-----VGSASVIYMIADFWGLVEENEVEHVGSS-- 941
D 465 MFKALERMPCLTLPVLSRLEEKNEWKS VKGS-----LQPCWRSIEFKNYSKLSLDSQCV 519
QY 942 --KFRDEE--ESFTLLSDCDVLVARS PAHFPSDIOQRVAVFPKPELHSLKDVILPSTKGV 997
D 520 YFKARDKQVSSKFLAEADILRSQAKLHFP--LRSRSAFEFSFYVDNEIVLFDT---- 572
QY 998 PLAKKLSGGDYDGMWVCMWDPDPIVDGFVNAEMPLEDLSRYLKDKTKTFKQLMASHGTG 1057
D 573 -----CY---MVCTIVCNSP-----SGLKKVEHFFKNILPLHFG 605
QY 1058 SAAKEQTTVDIMIQKGFHFALQ--PNFLGMCNTYKERLCYINNSVSNKPAIILSSLVGNLVD 1116
D 606 EXDKFSIFLDQVFRGPDYDVNAPNIVG-----NKFVRRKRSNSITQLT-EFYK 652
QY 1117 QSK-QGIVFNEASWAQLRREILGALSIPDPMYKSDSWLGRGEPHTIIDYLFKSIARPAI 1175
D 653 QPKINGQSRSAARAARKEESGNKSQNSQNSLSD-----ESGNVTPVSKQLSQPAA 706
QY 1176 DXELEAFHNAKMA-----AKDTGGAHFWDPD-----LASYTFPKEI 1213
D 707 -----AIKASLKYPSPHSDLSLEHODHAGDTENEMHD-DVDKEQFGYSVYVFFR-- 754
QY 1214 SDKSRSSALLFTLKNRIGEVK--EYGLVKNKEMRDSKDPYPVRVNVQV--EKWCAIT 1269
D 755 -----LFNLLYERLYELQRLSDQVSIQQRII-----PNPVSQKQKINRDRNDLS 800
QY 1270 PEAMDKSGANYDSKVIRLLELSF-LADREM--NTWALLRASTAPFKLYYHKSPKFWQWAG 1326
D 801 DVPDEKT--HYENTYVMILRIYIGIVDSAPEDYLRFYGNKAYKIY--TIDKLVWS-AA 855
QY 1327 RQLAYIKAQMTSRPGEAPALMTAFM--YAGLMPDKKTKQYVARELGDGSEYDPPEVYE 1384
D 856 KQVHHIVS-----DGKYKFTVSLVQNSASPKKNY-DDFLYRLEIEKLNLPDEILFR 907
QY 1385 VLGDGDDFDGIG 1395
D 908 FCWINKFKSFG 918

RESULT 9
F84582
Hypothetical protein At2g19910 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84582
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, I.
eure, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84582
A>Status: preliminary
A:Molecule type: DNA
```

A;Residues: 1-966 <STO>
A;Cross-references: GB:AE002093; NID:g3687227; PIDN:AAC62125.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g19910
A;Map position: 2

Query Match 2.4%; Score 177.5; DB 2; Length 966;
Best Local Similarity 19.8%; Pred.No. 0.014;
Matches 167; Conservative 116; Mismatches 267; Indels 233; Gaps 41;

QY 518 LMFEGGCRLTRRFGDRFPEILLIPSTSTSPSPVPPVSKQPGAVEEVIQWLTMGQHSILVG 577
DB 232 LLENTGTHLKVGLGDDNVLTKFEDKVLGVETVYCNLDLYTYGIAKNGI-----MVG 282
QY 578 -QWRAFFAKDAGYKPLREDFOLRAEDPKPIKERVH-FFAETGTTF-----RDPVFTR 630
DB 283 LRRYFFVFKDGG-----KEKKKDVSTGKVCYFIRTDSTASIDMQNPYIFAGK 332
QY 631 SVVPAEEPVEQTEFKVQMDLQDNTWQPHLKLFSRIQLGLSKTYAINTLBEHQI 690
DB 333 SM-----HEARHFM-----HNTLSLPNYSR-----YSI 359
QY 691 RHKTDLLSPSTGEVMNDGVGRMSVA-----KRIK-----DVLGLG-- 729
DB 360 QDDKDVLDKNGKPCIHSDGTGYISEDLARMCVPVNI PKGKSMRSNNIQSKNLNFEQGQPC 419
QY 730 --DVPSAVQGRP-----GSAKGWVLDVDDTGDDEWIEYTPQKRWCEDFVDKHQRTLEVR 783
DB 420 GQEPFLQLQFRIFYNGYAVKG-----TFLTNKPL-----PRTVQVR 456
QY 784 SVASELSKAGLNLQLPVLVEDRADRKVKVQRAIGDRINDLQ---RQFSEQKHALNPVE 840
DB 457 P-----SMIKVVEDKTLNLS-----TNSLEVVTSPNPKARLSR--- 493
QY 841 FRQVYESYSRATRVSGRVP---FLAGLPDSQBEETNLF-----MNSGFD--- 884
DB 494 -----NLVALLSYGGVENDPFLNLRNLTLEESKTIIFYSRAAPKAAINTGDDQYT 543
QY 885 -----PKOKYLQDIADWLQKCKDTLSKLNIRVGRSAYIYMIADFVGLVLENEV- 935
DB 544 ADMILVGLPLDFPYLKDRLSYLLKTERNALKAG-RFPIDESYIYINGTVDPTGELKENEIC 602
QY 936 ---HVGFSKFRDEBESFTLLSDCDVLVARSPA-HFPDSIORVRAVFKPELH---SLKD 987
DB 603 VILHSGQISG-----DVLVYRNPGLHF-GDIHVLKATYVKALEDVYVNAKF 647
QY 988 VIIESTKGDVPLAKLISGDDYDGMWVWCWQPEIYDGFVNAEMPLEPDLRYLKKDKTTF 1047
DB 648 AVFFPKGPRSLGDIAGDPGDMYFISRNPKLLEHF-----KPEPWWSS--SKPSKIY 701
QY 1048 KQLMASHGTGSAKEQTTYDMTQKSFHFALQFNFLGMCNTYKERLCYINNSVSNKPAIIL 1107
DB 702 -----CGRKPSLSLEELFEMFL-----KAPFC----- 728
QY 1108 SSLVGNLVDQSGQIVFNEASWAQIRRELLGALSPLDPMYKSDSWLGRGEPTHII--DY 1165
DB 729 -----KRDVIGNA-----ADCLGIMDPFTLGDDES 754
QY 1166 LKFSIARPA-IDKELEAFHNAKAADTEDGAHF-WDPDL-ASYVTFPKETISDKS--RSS 1220
DB 755 AXEKYERKKNLKLDIYYDALDAPK--KGAKVDLPDLEIKNPFHYVYERDPKQDPEST 811
QY 1221 ALLFTYLNRIQEVEKEGRLVKNEMRDSKDPYVVRNOVYKWCATPEA-WDKSGAN 1279
DB 812 SIL-----GLIFDTYVDSHNAEPPPSSEISKLWYFEDEPVPKSHMDKFTSW 856
QY 1280 YDS 1282
DB 857 YEN 859

RESULT 10
S15053

hypothetical protein YCR033w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YCR592
C;Species: Saccharomyces cerevisiae
C;Date: 07-May-1993 #sequence.revision 07-May-1993 #text_change 19-Apr-2002
A;Accession: S15053; S19445; S40918
R;Jia, Y.; Slonimski, P.P.; Herbert, C.J.
Yeast 7, 413-424, 1991
A;Title: The complete sequence of the unit YCR59, situated between CRY1 and MAT, reveal
A;Reference number: S15052; MUID:91335897; PMID:1872032
A;Accession: S15053
A;Molecule type: DNA
A;Residues: 1-1226 <JIA>
A;Cross-references: EMBL:X59075; NID:g4805; PIDN:CAA41799.1; PID:g4807
R;Herbert, C.J.; Jia, Y.; Slonimski, P.P.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19445
A;Accession: S19445
A;Molecule type: DNA
A;Residues: 1-1226 <HER>
A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42300.1; PID:e264374; PID:g1907
R;Wickstead, B.L.; Roberts, A.B.; Segliocco, F.A.; Brown, A.J.P.
Yeast 7, 761-772, 1991
A;Title: The complete sequence of a 7.5 kb region of chromosome III from Saccharomyces
A;Reference number: S40917; MUID:92133166; PMID:1776366
A;Accession: S40918
A;Molecule type: DNA
A;Residues: 1-304, 'A', 306-374, 'S', 376-403, 'Q', 405-434, 'V', 436-441, 'N', 443-481, 'E', 483-
>
A;Cross-references: EMBL:S78634; NID:g244237; PIDN:AAB21259.1; PID:g244239
C;Genetics:
A;Cross-references: SGD:S0000629
A;Map position: 3R

Query Match 2.4%; Score 174; DB 2; Length 1226;
Best Local Similarity 18.4%; Pred.No. 0.033;
Matches 218; Conservative 178; Mismatches 453; Indels 338; Gaps 52;

QY 84 KAASNNVVKAHADPTLPMSKEPPRAATAGQWALQTVLLEVMFMPNNPTGRTG 143
DB 38 KSSNNGFVSPPTADNSTNP-SVTPSTAS-----VPLETAAPGSTFG 77
QY 144 -----RTLGPGLSRLPTSTNTRKDEPANVPADPKESLTSATGPIHGAAIPLKFP 198
DB 78 IEAPRESRVDPSVSPSSSVSTRKIGSRYPDPVERSSSTTSFESMNTSTITHNT 137
QY 199 DPVNTGSKPSLSENLNCTKRAKGLSDNV-----AAAAAPPVPIASALDKVPTTRHA 253
DB 138 DIGNSYSRKTMSRYN-----POSTSTNVTHTFSSALSNAPPFYVANGSSRRP--RSM 188
QY 254 NTRDPTATGHRRAQDVSDFTSQGTSGS-----SVFSACR-----HNQSTTQSSPE 300
DB 189 DDYSPDVTNKLSTNNVSSVNNNSPHSYYSRKNKWSIGTSPRPFPDNDHVGNNMTTSNTS 248
QY 301 APFSQPREKEPVDATVPEAGHLIESPKGRTTKSHDNO-----PLSSSSQGETSFS 352
DB 249 IHOREPFWKAN-STTILKSTHSQSSPSL-HTKKFHDANKLDKPEASVKVETPSKDETKII 306
QY 353 TYYE-SFPSSGGGEGALPEPSRNGLARSEARSQVQVHAPVVAARLNINPKFPKWL-- 409
DB 307 SYHDNNFPKPSVKSNAPLEPDNIKVGEDALGKKEVHKG-----REIAKEHPTFVKM 361
QY 410 --HEAPLAVAWETRLFMHCKVDLEDESLGLKYPDSWSTARDV---TDIWKLYRLDAPR 464
DB 362 KEHDELEAKKVNKI-----NIDGKQDEIWTAKTVASAVEVSKESHK----- 405
QY 465 GKPFPEKPNPDVFTAMTGNFESKSAVLASVLDYNPNDSPTAPLYLV-----KLKP 517
DB 406 -----ELTRSVKESPEIRDYRAYDPKALKTDATKLTAVDDDNKSVEEP 450
QY 518 LMFEGGCRLTRRFGDRFPEILLIPSTSTSPSPVPPVSKQPGAVEEVIQWLTMGQ--HSL 575
DB 451 LEKVEG-----IPPLPKA-----ETRLWELKNQVRNKI 479

QY 576 VGRWRRAFAKAGYKPLRQLRAEDPKPIIKERVHFFAETGITPRDPVFKTRSVVPA 635
DB 480 ISKQKYL-----KKAIRN-----SEYFFAQNKLHQQA-TGLI-----ITKILSKKK 524
QY 636 BEPVEQRTFVKVQSLDML-LQLDNNNTWQPHLKFSRIQLGSKTYAIAWTLPEHQIRHHK 694
DB 525 EEHLK-----KINKHYPDLQ-----KYEKECEIITLSENLRKEEIEIENKEHELMEOK 576
QY 695 TLLSPSGTGEVNDGVRGMSVAVAKRIRDLVGLGDVPSAVQGRGSAKGMWVIDVDTG 754
DB 577 R-----RBEIG-----ETEKEKSLRH-----PSS----- 595
QY 755 DEDWIEYPSQKWECDYFQHQRTLEVRVASELSKAGNLQLLPLVLEDRAR----- 807
DB 596 -----SSSSRRNRADFYDDAEMENVLLQIDPNKYQAAATIPPLIDPIRKYSYKFC 649
QY 808 -----DKYKMQAIGDLNDLQRFQSEOKHALNPVEFRQWYVESYSSRATRVSHGVPFL 864
DB 650 DVNNLVTDKLWASRLKADSNFTDHEHSL-----FLEG-YLIHPKKFGKISH-----YM 699
QY 865 AGLPDSQEBTLPNLMNSGDPKQKYLQDIAMDLOKCDITLKSKLNIIVRSAYIYMA 924
DB 700 GGLRSEPECVLAH-----YRTKTYNYKOLLIDKNKRKMSAAAKRKKERS----- 747
QY 925 DFMGVLENEVHVGFSSKFR-----DEESF-----TLLSDCV----- 958
DB 748 -----NDBEVEDESKEESTNTIDKEKSENNAENVPVLVQGVSEVKGDPFLGPEKVE 801
QY 959 -LVARSFAPFPSPDIQVRVAFKPLHSLKDV-----IFSTKGDVPLAKKLGGDYD 1009
DB 802 NMTEKGEFAGLENAERV-----NDLKEAHEDEESNKSSVIETNNEVQIAP-KGGYRN 858
QY 1010 GDMAWCWDPEIVDGFVNAEMPLEDLSRYLKDKTTFFKQLMASHGTSAAKEQTYDNI 1069
DB 859 G-----YYPEETK-----ELDFSLNALQR-----KHK-----SAPEHKTSYMSV 894
QY 1070 QKSPHF-----ALQPNFLGCTNYKRLCYINNSVSNKPAIILSSLVGNLVDQS 1118
DB 895 RESQLPELKEFGSOWSLSEKLGTKSTWYVYQRNAANGWLIVDE-----TDLK 949
QY 1119 KQIVFNEASWAQ-----REELGGALSPLDPMYKSDSWLGRGEPTH 1161
DB 950 RDGTSSESQOILQIOPERNINAYSNIP-POORPALGVFVGQPTH 995

RESULT 11
G84582
hypothetical protein At2g19920 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-905 <S70>
A:Cross-references: GB:AB002093; NID:g3687226; PIDN:AAC62124.1; GSPDB:GN00139
C:Genetics:
A:Map position: 2

Query Match 2.3%; Score 169.5; DB 2; Length 905;
Best Local Similarity 21.6%; Pred. No. 0.038;
Matches 134; Conservative 100; Mismatches 209; Indels 178; Gaps 34;
QY 575 LVG-RQWRAPFAKAGYKPLRQLRAEDPKPIIKERVHFFAETGITPRDPVFKTRSV 632
DB 285 MVGLRRYQFFVFKDGG-----KAEKKRNSTQKVCYFTRTGSTASSDM----- 328

QY 633 VPAAEPV-----EQRTEFKVSQMLDWLLQDNNNTWQPHLKFSRIQLGSKT----- 679
DB 329 ---ENPYILSGVSIHEARMHF-----MHVHTLPSANYMARFSLILSKTKLEV 374
QY 680 --YAINTLPEHQIRH---KTDLLSPSGTGEVNDGVRGMSVAVAK-----RI 722
DB 375 DMTEITVMQIDDDIHCDSQNNVDLDKNKPRTHSDGTGYISDELARMCPNIFKGMKMS 434
QY 723 RDVLG-LGDVPSAVQGRF-----GSA-KGMWVIDVDDTGDDEWIEYPSQKWECDVFDKH 776
DB 435 NNTQGTVCQVEPPLLIQIRMFNDGSAVKGIFLL--NKNLPPQTVQVRPSMIK---VYKDKN 489
QY 777 Q-----RTLVRVASELSKAGNLQLLPLVLEDRARDKVRQOATG-----DELIN 822
DB 490 LSNFSTFNSLEVVTTSPNPPKRAKSLKNVALL-----SYGGVNDFFLDILLN 537
QY 823 DLQORSEQHAKLNRPFVFRQWYVESYSSRATRVSHGVRFFLAGLPDSQEBTLPNLMNSG 882
DB 538 TLEK-----KKTIP--FKVRAA-----GKAALHYGNMD-DKNAQLQIMM-AG 574
QY 883 FDPKKQKYLQDIAMDLOKCDITLKSKLNIIVRSAYIYMAIDFWGVLNEVHVGFSSK 942
DB 575 I-PLDEPPLKHYLSKLLKLEKDDKAG-KLPIDESYILMGTVDPTGELKEDEV----- 625
QY 943 FRDEESFTLLSDCVLVAESPA-HPSDIQVRVAFKPLH-----SLKDVIIFFSTKGDV 997
DB 626 -----CVIL-----NPGIHF-GDIHLKATYVKSLEQYVGVNSKYGVFFPKGPR 668
QY 998 PLAKKLGGDYDGDMAVWCWDPEIVDGFVNAEMPLEDLSRYLKDKTTFFKQLMASHGTC 1057
DB 669 SLGDEIAGGDFDGDVYFISRNPKLEHY-----KPSEPVWVSSSPRSKIYTRQ--PSELSP 722
QY 1058 SAAKEQTYDNIQKSFH-----PALQPNFLGM-CTNYKRLCYINNSVSNKPA 1104
DB 723 EQEEELFKIFLTKGTFSPSSVQIGAADSWLAIMDRFLTIGDENVKK-----AENKKKM 776
QY 1105 IILSSLVGNLVDQSKQIVFN 1125
DB 777 LKLDIYVDAIDAPKTGTEVN 797

RESULT 12
A55575
ankyrin 3, long splice form - human
N:Alternate names: ankyrin G
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change 20-Sep-1999
R:Kordeli, E.; Lambert, S.; Bennett, V.
J. Biol. Chem. 270, 2352-2359, 1995
A:Title: Ankyrin-G. A new ankyrin gene with neural-specific isoforms localized at the
A:Reference number: A55575; MUID:95138209; PMID:7836469
A:Accession: A55575
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-4377 <KOR>
A:Cross-references: GB:U13616; NID:g608024; PIDN:AAA64834.1; PID:g608025
C:Genetics:
A:Gene: GDB:ANK3
A:Cross-references: GDB:424503; OMIM:600465
A:Map position: 10q21-10q21
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
C:Keywords: alternative splicing; peripheral membrane protein
F:106-138/Domain: ankyrin repeat homology <AN01>
F:173-105/Domain: ankyrin repeat homology <AN02>
F:139-171/Domain: ankyrin repeat homology <AN03>
F:172-200/Domain: ankyrin repeat homology <AN04>
F:201-233/Domain: ankyrin repeat homology <AN05>
F:234-266/Domain: ankyrin repeat homology <AN06>
F:267-299/Domain: ankyrin repeat homology <AN07>
F:300-332/Domain: ankyrin repeat homology <AN08>
F:333-365/Domain: ankyrin repeat homology <AN09>

F:366-398/Domain: ankyrin repeat homology <AN10>
 F:399-431/Domain: ankyrin repeat homology <AN11>
 F:432-464/Domain: ankyrin repeat homology <AN12>
 F:465-497/Domain: ankyrin repeat homology <AN13>
 F:498-530/Domain: ankyrin repeat homology <AN14>
 F:531-563/Domain: ankyrin repeat homology <AN15>
 F:564-596/Domain: ankyrin repeat homology <AN16>
 F:597-629/Domain: ankyrin repeat homology <AN17>
 F:630-662/Domain: ankyrin repeat homology <AN18>
 F:663-695/Domain: ankyrin repeat homology <AN19>
 F:696-728/Domain: ankyrin repeat homology <AN20>
 F:729-761/Domain: ankyrin repeat homology <AN21>
 F:762-794/Domain: ankyrin repeat homology <AN22>
 F:795-827/Domain: ankyrin repeat homology <AN23>

Query Match 2.2%; Score 163.5; DB 2; Length 4377;
 Best Local Similarity 17.3%; Pred. No. 1.2;
 Matches 275; Conservative 205; Mismatches 538; Indels 575; Gaps 72;

QY 30 CVADTTLTPHRRKELAESDEDFGRDKIYRALNELYWRKDDSLNCAEAFIEAKAASN 89
 DB 1430 CNLNTLPAKKKETESDQDETEKTRQSFASLALRKRYSLYLPFG---MIERSTGATR 1486
 QY 90 WYPKAHA-----DPDTLPW-----SKEPPRAATAGQOWALQTV 122
 DB 1487 SLPTTYSYKPFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSNTSPASPLKS-MSVST- 1545
 QY 123 LLEVLRNRPMP-NNTPGRTFGRTLSGSGLSRPTSTNTKXDEPANVTFADPPKSLTR 181
 DB 1546 -----PSPIKSTLGASTSSVKSISDVASPI-RSLRTWSSPIKTVWSQSPYNIQVS 1595
 QY 182 SAT-----GPIHGAAILPKF---PDYNTGSK-----RFSLESENLNOCT 219
 DB 1596 SGTLABAPVTEATPLKGLASNTSSRTSPVTTAGSLLESRTITMTTPASPKNINMYS 1655
 QY 220 KRAKGLSNVAAAAAPPV-----PIASALD-----KVPTRRHANTRDPTATGHR 265
 DB 1656 SSLPFX---SIITSAAPLSSPLKSVSEPKSRVIVISAKITMASSLSPPVQMPGHAE 1712
 QY 266 ADQVDSFDTSQGTYSYSSVFSACRHNOSTQSSFEAPSPQPREKRPVDATVEAGHLIES 325
 DB 1713 VALNGSISPLKYASSTTLNCKATATLOEKISATNS-----VSVVSAATDTVEK 1765
 QY 325 PKGRTT-----KSHIDNOPSSSSQGTSTSTYIEFPFSS----- 361
 DB 1766 VFSTTTAMPFPLRSVVAAPAFQSLRTPSASALYTSLGSSISATTSSTVSSITTPVY 1825
 QY 362 GGECAIPEPSRNGLARSEARSQVQVHAPVVAARLRNIWFKPKWLHEAPLAVAEVT 421
 DB 1826 SVVNVLPFALKK-LPDSNSFTKSAALUSPIKTLTTET-----HPQP---HFSRT 1872
 QY 422 RLFMHCKVLEDESGLKYDPSWSTA-----RDVTDIWKTLYRLDAF-----RGKPFPEK 471
 DB 1873 SSPVKSSSLFLAPSALKLSTPSSLSQCEILKDVAEWKEDLMRTAILQTDVPEEKFPQE 1932
 QY 472 PNDVFTVMTGNFESKGAIVLSAVLDYNPNSTAPLYLVKLPLMFEQCRLTR--- 528
 DB 1933 LPKE-----GRIDDEEPFKIVERKYE-----DLVKVSEILKKDVCVNDKNSP 1974
 QY 529 -----RFGPDRFFEILIPSTSTSPSPVVPVSKOPGAVEE-----VI 565
 DB 1975 KSPKSKDGHSPEDDWTFESSEIREARQQAASQSPSLPERVQVAKAASEKDYNLTKVI 2034
 QY 566 QWLT--MGQHSLVGRQWRAFFAKDAG-----YRKP---LREFQLR-----REDKP 606
 DB 2035 DYLNTNIGSSSLTNLKYKPEDAKKGEGQKRVLPALIALQBHLKMPASMRSTSTSEKE 2094
 QY 607 IIERVHFHAETGITPRPD-----VFKTRSVVPASEPVEQRTPEFVSQWL 651
 DB 2095 LCKWASDFGTDILSPDDFSQHDQDKSLSDSGPEITSEKTPSNP--QSAE----- 2145
 QY 652 DWLLQLDNNWPHLKLFSRIQLGSKTYAINTLEPHQIRHHKTDLL-----SPSGTGE 705

DB 2146 -----TTGPK-PLFHEVPI-----PPVITETRTVVHVIRSYDPSA--- 2180
 QY 706 VMNDGVRMSRSVAKIRDLVLGLGYDPSAQVGRFGSAGKMWIVDDTDGDEWIEYPSQ 765
 DB 2181 -----GDPV-----QTQPEEPVSPKPS 2198
 QY 766 RKNEDFVDKHQRTLEVRVASBELKSAGNLQLLPVLEDRARDKVKRQAGRLINDLQ 825
 DB 2199 TFME-----LEPKPTTSSIK-----EKVRAFO-----MK 2222
 QY 826 RQFSEQKH--ALNRPVEFRQWVYESSRATRVSHGRVFPFLAGLPDSQEBETLNF-----L 878
 DB 2223 ASSEEDDNRVLKSGMRVKE--ETHITTTTRMYVHSPFGGEGASERIEETMSVHDIMKA 2279
 QY 879 MNSGFPDKQ-----KYLQDIANDLOKRCDT-----LKSKLNRVGRSAIYMI 923
 DB 2280 FQSGRDPFSKLAGLFEHKSAVSPDVHKSAEASQAHAEKDNOMKPKLERII----- 2330
 QY 924 ADFWGLEENEHVHGFSSKPRDEESFTLLSDCVLVARSPAFPSPIQKRVAVFKPELH 983
 DB 2331 -----EVHI-----EKGNAEPTVEIIRTKKH-----PE-- 2355
 QY 984 SLKDVIFSTKGVPLAKLKGDDYDGMVAVWCWDPEIVDGFVNAEMPLEPDLRYLKDD 1043
 DB 2356 --KEMVYQ-----KOLSRGDI-----NLKDLFPEK 2379
 QY 1044 KTTFKQLMASHGTGSAAKEQTTVDMIQKSFHPALQPNFLGMCTNYKRLCYINNSVSNKP 1103
 DB 2380 HDAPP--CSEEQQOEERELTAE-----ESLP 2404
 QY 1104 AITLSLVGNLVDQSGQIVFNEASWAQRLRELLGALSPLDPMYKSDSLWGEPTTHII 1163
 DB 2405 SYLESRRVNTPVSEED---SRPSSAQL-----ISDSYKTLKLSQ---HSI 2446
 QY 1164 DY-----LAFSIARPAIDKELEAFNAKAAKDTEDGA--HFWDPLASYTF 1209
 DB 2447 EYHDDSELSELRGSYRFA-EKMLLSKDLVSHS-----DTESVTDHAGPPS----- 2492
 QY 1210 FKEI--SDKSRSSALLFTLKNRIGVEKEYGRLVKNKMRDSDKPYVVRVNV----- 1261
 DB 2493 -SELOGSDK-RSREKIATAPKEI--LSKIYKDVSENGVGVKSKDEHFDKVTVLHYSNV 2548
 QY 1262 ----YEWCAITPEAMDKSGAN--YDSKYTR-----LLELS-FLADR-----EMNT 1300
 DB 2549 SSPKHAMWMTEDRLDRGREKLIYEDRVDRTVKEABEKLTEVQSPFRDKTEKLNDELQS 2608
 QY 1301 WALLRASTAFKLYHKSPK-----FVWQAGRLAYIKAAQMTSRPGE-----APA 1346
 DB 2609 PEKKARPKNGKEYSSQSPTSSSPKVLTLTELLASNDENVKARQHGPQGGFPKAEKAPS 2668
 QY 1347 LMTAFMTAGLMPDKKFTKYQV-----ARLEGDS 1375
 DB 2669 LPSS-----PEKMWLSQOTEDSKSTVEAKGS 2694

RESULT 13
 T30868
 RhoA-binding protein p116Rip - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30868
 R:Gebbink, M.F.B.G.; Poland, M.; Kranenburg, O.; VanHoeck, F.P.G.; Moolenaar, W.H.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z20911
 A:Accession: T30868
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1024 <GB>
 A:Cross-references: EMBL:U73200; NID:g1657836; PID:g1657837; PIDN:AAB18198.1
 A:Experimental source: brain
 C:Function:
 A:Description: may inhibit RhoA-stimulated contractility and promote neurite outgrowth

[illegible]

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 52 |
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|

QY 673 ---QLGUSKTVAIMTLEPHQIRHHKHTDLSPSGTVGEVMDGVGRMSRVAKRIR-DVLGL 728
Db : : : : :
572 VENKLSRQAQVNLTV-----DKPD--PPAGT-----PCASDIRSSSLT 610
QY 729 GDVPSAVOGRGSGAKMVIIDVDTGDSWTETVPSQRKWCDFDQKHQRTLEVRVASE 788
Db : : : : :
611 SWYSSYDG--GSAVQSVSVIWSVDTKWE-----LAT 643
QY 789 LKSAGLNLO-LLPVLEDRARDKVKMRQAIGDRLINDIQRQFSEQKHALNRPFVQWVYE 847
Db : : : : :
644 CRSTSFNVQDLLP-----DREYKFRVRAIN-----VYG 671
QY 848 SYSRATVSHGRVPFLAGLSDSQEETINFLMNSGFPKQKYLQDIANDLQKRCDTLK 907
Db : : : : :
672 --TSBPQSE--LTLGKPEE-----BPKDEVBVD-----DDEKE 705
QY 908 SKLIRVGRSAYIIMYADFVGLBEVHVHVSFKFRDEES-----FTLLSDCDV 958
Db : : : : :
706 PEVDYRTVTNTEQKVSDFYDI-EERLGSKGFGVFLVEKTKCIWAGKFFKAYS-----760
QY 959 LVASPSAHFPSDIOVRVAFKPELHSLKDVIIFTKGDVPLAKKL-SGGDYDGDMAWCV 1017
Db : : : : :
761 --AKEKENIROEISIMNCHHPKLVQCVD--PEKANIVMVLVIVSG-----805
QY 1018 DPEIVDGFVNAEMPL-BDLSRYLKKDKTTFKQLMASHGTGSAKEQTIVDMIQKSPHEA 1076
Db : : : : :
806 --ELFERIIDDFELTERECIKYMKQ-----ISEGVYIHKQIGIVHLD 846
QY 1077 LQPNFLGCTNYKERLCYVNNVSNKPAILLSSVGLMVDQSKOGIVNFBASNAQ-----LR 1133
Db : : : : :
847 LKPNII-MCV-----NKTETRIKLIDFGLARLENAAGSLKLVFGTFEFAVEVIN 895
QY 1134 RELGGGALLSLPDPMPKYSW-LGRGEPHTIDYLKFSIARPAI-DKLEAFHMAKAAKD 1191
Db : : : : :
896 YEPIG-----YATDMSIG-----VICYLVSGLSPFGMDNNEFLANVTSATWD 940
QY 1192 TEDGAHFWDPLASYTFYFFKEISDKRS--SALLFTTLKNRIGVE-----KEYGRL 1241
Db : : : : :
941 FDDEA-----FDEISDDAKDFISNLLKCKMKNRLNCTCLOHPWLMKDTFNM 987
QY 1242 VKNEMRSDKDPYPRVNVQVYKWCATPEAMDKSG--ANYDSKVITLLELSFLADREM 1299
Db : : : : :
988 EAKKLSKDRKKYMAR-----RKW-----QKTGNVAVRAIGLSMAWISGLSGKSS 1034
QY 1300 TWALLRATAFKLYHHKSPKRVQWQAGRLAYIKAQWTSRPGEGAPALMTAFVYAGLMPD 1359
Db : : : : :
1035 TGSPTSIPNAEKLSESDVSQAFLVAEEXPHVKP-----VFSTIRD 1077
QY 1360 KKFTKQVVARLEGDSYPPDEVY-----EVLGDDDD 1390
Db : : : : :
1078 LEVVEGSAARFDCKIEGYPDEVVWFKDQDSIRESRHFIQDYDEGNCSLIISDVCGDDD 1137

RESULT 15
H84582
hypoetical protein At2g19930 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C/Accession: H84582
M.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: H84582
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-929 <STC>
A/Cross-references: GB:AE002093; NID:G3687225; PIDN:AAC62123.1; GSPDB:GN00139
C/Genetics:
A/Map position: 2

| Query Match | 2.1%; Score 159; DB 2; Length 929; |
|---|---|
| Best Local Similarity | 19.2%; Prec. No. 0.17; |
| Matches 179; Conservative 140; Mismatches 318; Indels 293; Gaps 46; | |
| QY | 514 KKLPMFEQGCRLRRRQDPFRFBILLTPSTSPSPVVKQPGAVEEIOMLTGWQH 573 |
| DB | 193 RMGELLEHTGHLHKVLGDDNLTVKADYQKSSSTY-----SIDHYFYKGIAGN 243 |
| QY | 574 S-LVG-RQWRAFFAKDAGYKKPLREFOLRAEDPKPIIKERVH-FFAETGTFTRPDVPKTR 630 |
| DB | 244 GIMTGLRARYQFFVFPKGG-----KEKKDLSTKKVKCYFIRDTSTAFYDM-QNP 292 |
| QY | 631 SWPAEBPVEORTBF---KVSOMLDMLQLDNNTWQPHLKLPSRIQLGSKTVAIMTLE 686 |
| DB | 293 YILGKSIYEABMFHMHVRAPTLANYM-----ARPSLLISKT---KYLE 334 |
| QY | 687 P-----HQTRHKTDLSPSGTEVWNDGVGRMSRSVAKRIDVLGLGV-- 731 |
| DB | 335 VDMGTITFDQIDTHCHDQDGKVDKNKPCIHSDGTGYISDLARMCPNIFKKGCLR 394 |
| QY | 732 ----PSAVQGRF---GSAKGWVIDVDGTGDEDMTETYPQKWECDFVDKHORTLEV 782 |
| DB | 395 SESTQPLLIQRFMYDGYAVKGTFLN-----KKLC-----PRIVQV 431 |
| QY | 783 R-SVASSELKAGLN-----LQLLVLEDRDKVKMRQAIGRLIN--DLQRQFSQKHAL 835 |
| DB | 432 RPSMIKVSQKPSLNSFTFNALEYVITRS-----QIFGAAALNYGEMDQNAQAQM--- 481 |
| QY | 836 NRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEBTLFNLMSGDFPKKOKYLQDIA 895 |
| DB | 482 ----LKTENDLKAGKL--PVYESYLMGTVDPTGALKEDDEVCLTESQISGE----- 548 |
| QY | 956 CDVLVARSRA-HRPSDIOVRVAFKPELH---SLKDVLIFSTKGDVPLAKLSSGDYDG 1010 |
| DB | 501 --LKTENDLKAGKL--PVYESYLMGTVDPTGALKEDDEVCLTESQISGE----- 548 |
| QY | 549 --VLVYRNPGLHP-GDIHLKATYVALEBYVGNSTFAVFFPKGGRSLGDETAGDFFDG 605 |
| QY | 1011 DMAVWCWDPPIVQGVNAEMPLEPDLRSYLKKDKTTFKQLMASHGTSAAKEQTTVDMIQ 1070 |
| DB | 606 DMVFISENPBLENF---KPSFPMWS-----LTPPSKSNSGRAPSQLSPBEELE 651 |
| QY | 1071 KSHFHALQPNF---LGMCTN---YKBRCLVINNNSVSNKPAIILSSVLGNLVDQSKOG 1121 |
| DB | 652 ELEMFTATAGFHASNVIGIADSWLITMDRFLJLGDRAEERKEMKKML-ELID----- 705 |
| QY | 1122 IVRNEASWAQLRRELJGALSPL-----DPMYKSDSWLGRGEPTHIIDYK 1167 |
| DB | 706 -IYYDALDAPKK---GDKVYLPNKLKPDIFPHYMERDKKFQSTSLGL-----IDFVK 755 |
| QY | 1168 FSIAR---PAIDKELEAFHNAKAAKDTEDGAHFWDPDLASVYTFPKETSDKSRSSALLF 1224 |
| DB | 756 SQTEPSPSSGK-----GPLLVLVLT-YNQTKEIINHDMCF 792 |
| QY | 1225 TTLKNRIGEVK-----EYGRLVKNKEMRSDKDPYPRVNVQVYE 1263 |
| DB | 793 TVVYNTL--EISKLPCFEDEBPSEFFHMQKRLVYDNYRTWMTQAMKTDKDESANEVIQRYK 850 |
| QY | 1264 KWCAITPEAMDKGSANYDSKVIHLLLELSFLADREMNMTWALLRASTAFKLYH---KSPYF 1320 |
| DB | 851 Q-----EFVGAAGFE-DSK--KSLB-----ELYPQALALYKIVDYVAIHAGVSKCRF 894 |
| QY | 1321 VWQWAGRQLA--YIKAQMTSRPGEAGAPALM 1348 |
| DB | 895 VWKACQGVLCRFYLNKKMQEKKIIVCAPSVL 924 |

Search completed: March 29, 2004, 06:50:00
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:21:11 ; Search time 19 seconds
(without alignments)
3842.229 Million cell updates/sec

Title: US-09-913-878A-2

Perfect score: 7397

Sequence: 1 MNPITPRKNSPVEIINRL.....YEVLGDDDFDGIQGTNGDY 1402

Scoring table: BLOSUM62

Searched: i41681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 182 | 2.5 | 1075 | 1 | PST2_SCHPO |
| 2 | 168 | 2.3 | 1226 | 1 | SNT1_YEAST |
| 3 | 163.5 | 2.2 | 4377 | 1 | ANK3_HUMAN |
| 4 | 160.5 | 2.2 | 1024 | 1 | RP3_MOUSE |
| 5 | 160.5 | 2.2 | 5147 | 1 | PCLO_HUMAN |
| 6 | 160 | 2.2 | 1176 | 1 | KMLS_BOVIN |
| 7 | 158.5 | 2.1 | 1736 | 1 | ZOI_HUMAN |
| 8 | 157 | 2.1 | 1026 | 1 | TAC2_HUMAN |
| 9 | 156 | 2.1 | 2067 | 1 | NCO6_MOUSE |
| 10 | 154.5 | 2.1 | 5120 | 1 | PCLO_CHICK |
| 11 | 153 | 2.1 | 5430 | 1 | NACF_HUMAN |
| 12 | 153 | 2.1 | 5938 | 1 | NAC4_HUMAN |
| 13 | 151 | 2.0 | 1035 | 1 | TAC2_MOUSE |
| 14 | 151 | 2.0 | 3644 | 1 | MINT_MOUSE |
| 15 | 148.5 | 2.0 | 1418 | 1 | SNC4_YEAST |
| 16 | 147.5 | 2.0 | 1385 | 1 | FAT1_SCHPO |
| 17 | 144 | 1.9 | 980 | 1 | RIN3_MOUSE |
| 18 | 144 | 1.9 | 2224 | 1 | PA5_HUMAN |
| 19 | 143.5 | 1.9 | 848 | 1 | YBV8_YEAST |
| 20 | 143.5 | 1.9 | 860 | 1 | CH12_COCHO |
| 21 | 141.5 | 1.9 | 3660 | 1 | DMD_CHICK |
| 22 | 140.5 | 1.9 | 1516 | 1 | Y819_PSEAE |
| 23 | 140.5 | 1.9 | 3587 | 1 | SRF2_BACSU |
| 24 | 140 | 1.9 | 1001 | 1 | PSPA_RAT |
| 25 | 139.5 | 1.9 | 1367 | 1 | AMVH_YEAST |
| 26 | 139 | 1.9 | 3685 | 1 | DMD_HUMAN |
| 27 | 137 | 1.9 | 5171 | 1 | BPEA_HUMAN |
| 28 | 136.5 | 1.8 | 1403 | 1 | YDF3_SCHPO |
| 29 | 136 | 1.8 | 1273 | 1 | WEB1_YEAST |
| 30 | 135.5 | 1.8 | 956 | 1 | YEF3_YEAST |
| 31 | 135.5 | 1.8 | 3969 | 1 | HRX_HUMAN |
| 32 | 135 | 1.8 | 836 | 1 | NOT3_YEAST |
| 33 | 135 | 1.8 | 877 | 1 | WEB1_SCHPO |

| | | | | | |
|----|-------|-----|------|---|------------|
| 34 | 135 | 1.8 | 3396 | 1 | PGCV_HUMAN |
| 35 | 134.5 | 1.8 | 3588 | 1 | SRF1_BACSU |
| 36 | 133.5 | 1.8 | 1723 | 1 | AIM1_HUMAN |
| 37 | 133.5 | 1.8 | 1967 | 1 | YGSO_YEAST |
| 38 | 133.5 | 1.8 | 2150 | 1 | SDC3_CAEEL |
| 39 | 133 | 1.8 | 1487 | 1 | ICP4_HSVBK |
| 40 | 132.5 | 1.8 | 3343 | 1 | BRC2_RAT |
| 41 | 132 | 1.8 | 1395 | 1 | IFAG_HUMAN |
| 42 | 132 | 1.8 | 1953 | 1 | BN11_YEAST |
| 43 | 132 | 1.8 | 2418 | 1 | SPCA_HUMAN |
| 44 | 132 | 1.8 | 3331 | 1 | ESYN_FUSEQ |
| 45 | 131.5 | 1.8 | 1337 | 1 | DEXT_STRDO |

ALIGNMENTS

| | | | | | |
|------------|--|-----------------------------------|------|----------|--|
| RESULT 1 | | | | | |
| PST2_SCHPO | | | | | |
| ID | PST2_SCHPO | STANDARD; | PRT; | 1075 AA. | |
| AC | O13919; | | | | |
| DT | 16-OCT-2001 | (Rel. 40, Created) | | | |
| DT | 16-OCT-2001 | (Rel. 40, Last sequence update) | | | |
| DT | 10-OCT-2003 | (Rel. 42, Last annotation update) | | | |
| DE | Paired amphipathic helix protein pst2 (SIN3 homolog 2). | | | | |
| GN | PST2 OR SPAC23C11.15. | | | | |
| OS | Schizosaccharomyces pombe (Fission yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; | | | | |
| OC | Schizosaccharomycetales; Schizosaccharomycetaceae; | | | | |
| OC | Schizosaccharomyces. | | | | |
| OX | NCBI_TaxID=4896; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=972; | | | | |
| RX | MEDLINE=21848401; PubMed=11859360; | | | | |
| RA | Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A., | | | | |
| RA | Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S., | | | | |
| RA | Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., | | | | |
| RA | Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., | | | | |
| RA | Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., | | | | |
| RA | Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., | | | | |
| RA | James K., Jones L., Jones M., Leather S., McDonald S., McLean K., | | | | |
| RA | Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., | | | | |
| RA | Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E., | | | | |
| RA | Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., | | | | |
| RA | Skellton J., Simmonds M., Squares R., Squares S., Stevens K., | | | | |
| RA | Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., | | | | |
| RA | Woodward J., Volkart G., Aert R., Robben J., Grymonprez B., | | | | |
| RA | Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., | | | | |
| RA | Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., | | | | |
| RA | Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., | | | | |
| RA | Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B., | | | | |
| RA | Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., | | | | |
| RA | Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., | | | | |
| RA | Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G., | | | | |
| RA | Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., | | | | |
| RA | Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., | | | | |
| RA | Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., | | | | |
| RA | Shpakovski G.V., Ussery D., Barrell E.G., Nurse P., | | | | |
| RT | "The genome sequence of Schizosaccharomyces pombe." | | | | |
| RL | Nature 415:871-880(2002). | | | | |
| RN | [2] | | | | |
| RP | GENE NAME. | | | | |
| RX | MEDLINE=99147069; PubMed=10022921; | | | | |
| RA | Dang V.D., Benedik M.J., Ekwall K., Choi J., Allshire R.C., | | | | |
| RA | Levin H.L., | | | | |
| RT | "A new member of the Sin3 family of corepressors is essential for | | | | |
| RT | cell viability and required for retroelement propagation in fission | | | | |
| RT | yeast." | | | | |
| RL | Mol. Cell. Biol. 19:2351-2365(1999). | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE OF 87-104; 280-291; 386-395; 424-446; 508-637; 678-709; | | | | |
| RP | 780-788; 819-834; 869-883; 925-937 AND 1055-1075; FUNCTION, SUBUNIT, | | | | |

| | |
|--------|-------------|
| P13611 | homo sapien |
| P27206 | bacillus su |
| Q9Y4K1 | homo sapien |
| P53327 | saccharomyc |
| P34706 | caenorhabdi |
| P17473 | equine herp |
| O35923 | rattus norv |
| O04637 | homo sapien |
| P41832 | saccharomyc |
| P02549 | homo sapien |
| Q00869 | fusarium eq |
| P39653 | streptococc |

```

RP AND SUBCELLULAR LOCATION.
RX MEDLINE=2265054; PubMed=12773392;
RA Nakayama J.-I., Xiao G., Noma K.-I., Malikzay A., Bjerling P.,
RA Ekwall K., Kobayashi R., Grewal S.I.S.;
RT "Alp13, an MRG family protein, is a component of fission yeast Clr6
RT histone deacetylase required for genomic integrity.";
RL EMBL J 22:2776-2787(2003)
CC -!- FUNCTION: Has a role in chromatin assembly and chromosome
CC segregation. Involved in the deacetylation of histones.
CC -!- SUBUNIT: Heterotrimer of alp13, clr6, prw1 and pst2.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 3 PAH (paired amphipathic helix) repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z98559; CAB11171.1; -.
CC FIR; T38253; T38253.
CC GenesDB Spombe; SPAC23C11.15; -.
CC InterPro; IPR003822; PAH.
CC Pfam; PF02671; PAH; 3.
CC Transcription regulation; Chromatin regulator; Nuclear protein;
CC Repeat.
CC Repeat. 54 100 PAH 1.
CC Repeat. 160 206 PAH 2.
CC Repeat. 271 317 PAH 3.
CC SEQUENCE 1075 AA; 124848 MW; 1AD301DB4EB1AFFB CRC64;
CC
CC Query Match 2.5%; Score 182; DB 1; Length 1075;
CC Best Local Similarity 18.7%; Pred. No. 0.0061;
CC Matches 193; Conservative 157; Mismatches 359; Indels 322; Gaps 51;
CC
QY 494 LSAVLDPNDSPTAPLYL-VKLPLMFEQGRLLTRRGDPRFFELLIPS-PTSTSPSVP 551
DB 81 ISVILRDYDPLLEYLNIFLPSYKLLNSGANFTLQF-----TTPSGFVSTPTIV 132
QY 552 PVVSKPGAAVEEVIWLTGQSHLVGRQWRAFAKADAGYKPLRFQ----- 598
DB 133 ATYNDLPCTYHRAIGFVSVRALLSNPQFFKLQDS-----LRFKNSECSLSLQIV 187
QY 599 --LRADPKPIKERVHFAETGTF--RPDV--FKTSVVPAPFVQRTPEFKVQSLD 652
DB 188 TSLLABHPS--LAHEPHNPLSPISFFGSKPPLGSPFLRG-----QSSQFTLSNISD 237
QY 653 WLLQLDNTWQHLKLSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGVNMDGV 712
DB 238 LLSQ-----SEPDNLSP--FSLHLSNE--- 256
QY 713 RMSRSVAKIRDLV-----GLGVPSAVQGRFG---SAGKMW--VID 749
DB 257 --SSDFKQXNVLTDTVYHFLKLLNLYVQGIIDRNILVSRGFGFLKNSGLWRSFLS 314
QY 750 VDDTGDEDIETYPQSKWECDFVDKHQRTLEVRVASELKSAGLNQLLPVLEDRADK 809
DB 315 LTLSPSEELSVT-----NGACDFPCGFSYRLLPVEERNISCS 354
QY 810 VKMRQAIGRLIND-----LQR-----QFSEQHNLNRPVEFRQW 844
DB 355 GRDDFAWG--ILNDDVWSHTWASESGFIVQKTPYEBAMTKLEERYEFDRIEATSW 412
QY 845 VYESYSRATRVSHGRVFLAGLPDSQETLNLNLSGDFDPK-----QKYLQD 893
DB 413 TIKSLKKIONRINE-----LPEERETYLEGLGLPSKIYKTKLVTSEHAE 464
QY 894 IANDLQKRCDDTLK---SKINIR-----VGRSAVIYMIADFWGLENEVHVGFSS--- 941
DB 465 MFXALRMPCLTLPLVIRLEEKNEVKWSKGS-----LQPCWRISFPQVNDKSLDSCV 519

```

RESULT 2.

```

SNTL_YEAST STANDARD; PRT; 1226 AA.
ID SNTL_YEAST
AC SNTL_YEAST Q02397; Q0N1L8;
DT 01-MAY-1992 (Rel. 22, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Potential DNA-binding protein SNTL.
GN SNTL OR YCR033W OR YCR33W OR YCR592.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae;
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92133166; PubMed=1776366;
RA Wicksstead B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;
RT "The complete sequence of a 7.5 kb region of chromosome III from
RT Saccharomyces cerevisiae that lies between CRY1 and MAT.";
RL Yeast 7:761-772(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335897; PubMed=1872032;
RA Jia Y., Slonimski P.P., Herbert C.J.;
RT "The complete sequence of the unit YCR59, situated between CRY1 and
RT MAT, reveals two long open reading frames, which cover 91% of the
RT 10.1 kb segment.";
RL Yeast 7:413-424(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandrak D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Benit P., Berben G., Bergantino E., Bateau N., Bolle P.-A.,
RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,

```

RA Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA De Haan M., DeBoor E., Delgado M.D., Demolder J., Dolra C., Dubois E.,
RA Dujon B., Duesterhoef A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Francinques-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA Fuller L.J., Gent M.E., Gigot D., Gallicquet V., Glandsdorff N.,
RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
RA Harat D., Hegmann J.H., Herbert C.J., Hilger P., Hohmann S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Kleinhaus U., Kreis P., Lafranchi G., Lewis C., van der Linden C.G.,
RA Luchini G., Lutzenkirchen K., Maat C.T., Mannhaupt G., Manzano M.E.,
RA Martegani E., Mathieu A., Maurer C.T., McConnell D., McKee R.A.,
RA Messenguy F., Meves H.-W., Mollema F., Montague M.A., Navas L.,
RA Newlon C.S., Owens M.V., Pallier C., Panzeri L., Pearson B.M.,
RA Perea J., Philippsen P., Piller A., Planta R.J., Plevani P.,
RA Poetsch B., Pohl F.M., Purnelle B., Ramazani Rad M., Rasmussen S.W.,
RA Raynal A., Remacha M., Richters A., Roberts A.B., Rodriguez F.,
RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala J., Stoninski P.P., Sor F., Soustelle C.,
RA Spiegeberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA Thores G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
RA van Vliet-Kedijk J.C., Volckart G., Vreken P., Warmington J.R.,
RA Zimmermann F.K., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RT "The complete DNA sequence of yeast chromosome III.",
RL Nature 357:38-46(1992).
RN [4].
RP REVISIONS.
RA Valles G., Volckaerts G.;
RN Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RL [5]
RP IDENTIFICATION IN A COMPLEX WITH SET3; HST1; HOS2; SIF2; CPR1 AND
RP HOS4.
RX MEDLINE=21567937; PubMed=11711434;
RA Pijnappel W.M.P., Schaft D., Roguev A., Shevchenko A., Tekotte H.,
RA Wilk M., Rigaut G., Seraphin B., Aasland R., Stewart A.F.;
RT "The S. cerevisiae SET3 complex includes two histone deacetylases,
RT Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation
RT gene program.";
RL Genes Dev. 15:2991-3004(2001).
CC -!- FUNCTION: Part of the Set3C complex, which is required to repress
CC early/middle sporulation genes during meiosis.
CC -!- SUBUNIT: Identified in a Set3C complex with SET3, HST1, HOS2,
CC SIF2, CPR1 and HOS4.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 Myb-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; S78624; AAB21259.1; -;
CC EMBL; X59075; CA441799.1; -;
CC EMBL; X59720; CAC42983.1; -;
CC PIR; S15053; S15053.
CC GeneOnline; 138939; -;
CC SGD; S0000629; SNT1.
CC GO; GO:0000118; C:histone deacetylase complex; IDA.
CC GO; GO:0003677; F:DNA binding; NAS.
CC GO; GO:0037136; F:NAD-dependent histone deacetylase activity; IDA.
CC GO; GO:0045129; F:NAD-independent histone deacetylase activity; IDA.
CC GO; GO:0016575; F:histone deacetylation; IDA.
CC GO; GO:0045835; P:negative regulation of meiosis; IDA.
CC InterPro; IPR007137; DJF348.
CC InterPro; IPR001005; Myb_DNA_binding.
CC Pfam; PF03990; DJF348; 1_DNA_binding.
CC Pfam; PF00249; myb_DNA-binding; 2.
CC SMART; SM00711; SANT; 2.

DR PROSITE; PSS0090; MYB_3; 1.
KW Nuclear protein; DNA-binding; Coiled coil.
FT DOMAIN 28 159 SER-RICH.
FT FT 539 591 COILED COIL (POTENTIAL).
FT FT 594 634 POLY-SER.
FT FT 594 599 MYB.
FT FT 305 305 A -> T (IN REF. 2).
FT FT 375 375 S -> N (IN REF. 2).
FT FT 404 404 Q -> H (IN REF. 2).
FT FT 435 435 V -> A (IN REF. 2).
FT FT 442 442 N -> D (IN REF. 2).
FT FT 482 482 E -> K (IN REF. 1).
FT FT 513 513 L -> V (IN REF. 2).
FT FT 644 644 H -> Y (IN REF. 2).
FT FT 766 766 E -> D (IN REF. 2).
FT FT 806 806 Q -> K (IN REF. 2).
FT FT 831 831 V -> I (IN REF. 2).
FT FT 846 846 G -> E (IN REF. 2).
FT FT 851 851 D -> A (IN REF. 2).
FT FT 857 857 AVQ -> GVR (IN REF. 2).
SQ SEQUENCE 1226 AA; 138396 MW; 70AB662227088EB CRC64;
Query Match 2.3%; Score 168; DB 1; Length 1226;
Best Local Similarity 18.8%; Pred. No. 0.051;
Matches 221; Conservative 174; Mismatches 466; Indels 314; Gaps 50;
QY 84 KAASNNWPKAHADPTLPWSKEPPRAATAGQWALQTVLLEVLNRFMPPPNNTTGTGTG 143
DB 38 KSNNGFVSPTADNSTWP-SVTPSTAS-----VPLTAAAGSTEG 77
QY 144 -----RTLSGPGSLGRPTSTNTKRDDEPANVTADPPKSLTRSATGPIHGAALPLKEP 198
DB 78 IEAPRSRYDPSVSRRSSSYSTRKIGSRYPDVERSSSTSTSTPSMMNTTTHHTWT 137
QY 199 DPVNTGSKRPSLESENINQCTKRAKGKLSDNV-----AAAAAPPVPIASALDKVTRRHA 253
DB 138 DIGNSRYSKTMSRYN-----POSTSTNTWTFPSALSNAAPPVAVANGSSRRP-RSM 188
QY 254 NTRDPTATGHRADQVDSPTSCQTSYSG-----SVFSACR-----HNGSTQSFE 300
DB 189 DDYSPDVTNKTETNNVSSVNNNSPHYSYRSNKNRSIGTSPRPDPDHNHVNNTTSTNNS 248
QY 301 APPSQPREKRPVDATVFEAGHLIESPSKGRITKSHIDNQ-----PLSSSSQGETSFS 352
DB 249 IHCQEPFWAN-STYLSKTHSQSSPSL-HTKKFHDANKLDKPEASVKVEPSKDETKAI 306
QY 353 TYTE-SPPSSGGEAIPSPRSNGIARSESARSQVQVHAPVVAARLNINWPKFKWL-- 409
DB 307 SYHDNFPKPKSVKPNAPLEPDNIKVGEEDALGKKEVHSG-----REIAKEHPTPKM 361
QY 410 --HEAPLAVAWETRLFMWCKVDLESLGLKVDPSWSTARDVTDINKLYELDAFRGK 467
DB 362 KEHDELEAKKVKSKI-----NIDGQDEIWTAKTVA-----SAVEVSK 401
QY 468 FPEKPNDFVFTAMTGNFESKGSAYVLGAVLDYNDPNSPTAPLYLVKL-----KP 517
DB 402 ESQK-----ELTRSVRKESPEIRDYERAYDFKALKT-----VTKLTVNDNKSIEP 450
QY 518 LMFEQOCRLTRRFGDPRFEIIPSTSTSPVPVVKQPGAVEVQWLTMGQ--HSL 575
DB 451 LEKVEGC-----IFPLPKA-----ETRLWELKNQKRNKI 479
QY 576 VGRQWRAEFAKDAGYRKPLREPOLRAEDPKPIKERVHFFAETGITFRDPVFKTRSVPA 635
DB 480 ISEQKYL-----KKAIRNF---SEYFFYAQNKLHQQA-TGLI-----LTKIISKIK 524
QY 636 EBPVEORTFKVSMQMDWL-LQDNNWQPHLKLPSRIQLGLSKTYAIVTLBPHQIRHHK 594
DB 525 EBHLK-----KINLKHIDYDLO---KKYEKECEIITLSENLRKEIEIKRKEHELMQEK 576
QY 695 TDLLSPSGTGEVMDGVGRMSRAKIRDLVLGLGDPGSAVQGRFGSAGKGMVIDVDDTG 754
DB 577 R-----REGI-----ETEKESLRH-----PSS----- 595

| | | |
|---|------|---|
| Query Match | | 2.2%; Score 160.5; DB 1; Length 1024; |
| Best Local Similarity | | 19.9%; Pred. No. 0.11; |
| Matches 211; Conservative 153; Mismatches 407; Indels 291; Gaps 51; | | |
| QY | 161 | KKKDDPANVTFAADPKRSLTRSATGPIHGAIPKFPDPVNTGTSKRPSLESENINQCTK 220 |
| DB | 159 | KKKVEPTTQEPGPAKAVTSSGGTSGSSSIP-----SAEKVPTTKS-TLWQEM 209 |
| QY | 221 | RAKGKLSNDVAAAAPP-----VPIASALDKVPTRRHANTRDPTACHRADQVDFDTSQ 277 |
| DB | 210 | RAKDQPDGTSLSPAQSPSQSPAACTPREP-GLSKEDESTISG-----DRVDGGRK 261 |
| QY | 278 | TSYGSVFACRHNGSTQSSPEAP-----PSQPREKRPVDATVFEAGHILIESPKGRT 331 |
| DB | 262 | VRVESGYFSLERAKQDLRAEBEOLPILLPPSPSTPHSRSSQVIEKFEA-LDIEKAHME 319 |
| QY | 332 | TKSHIDNQPLSSSSQGET-----SPSTYIESFPSSGGEGALPEPSRNGLARBEES 382 |
| DB | 320 | TNNLITLTPSSDTROGRSERRAIPKRDFASEAPTAPLSDACPLSPHRRAKSLDRRSTES 379 |
| QY | 383 | ARSOQVHAPVVAARLNWPFPK-WLHEAPLAWBE-----VTRLFMHCKVDLEDES 435 |
| DB | 380 | S-----MTPDLLN-----FKGWLTKQVEDQWKHFWLADQSLRYRVSVAEEA 425 |
| QY | 436 | LGKVDPSSTARDVDIMKTLRLDAFGKPFPPKPNDFVMTANGFSGKSAVVLS 495 |
| DB | 426 | ADLDGEINLSTCYDYTE-----YPQVQNYGFIHTK-EGEFTLSANTSGIRRNWQITMK 479 |
| QY | 496 | AVLDYNDPNSP--TAPLXLVKLPLMFEGQCLTRFRFGDPDRFFELIFSPSTSPSPVPV 553 |
| DB | 480 | HVL---PASAPDVTSLSLPGKKKTSFETCSRSTK-----QEAREGEFDP 522 |
| QY | 554 | VSKQGAIVE-----EVIQWITMG--QHSVLGRQWRAPFAKADQYRKPL-----RE 596 |
| DB | 523 | EOKKSARERREGSKTDFWAEFRPIQALQAERASAVGSDSDGDPCLAEPEGLERE 582 |
| QY | 597 | FOLRAEDPK-----PIKE---RVHFEATGTFRDPVETKRSV-VPAEE--- 637 |
| DB | 583 | RARRREPRRGMGLTDIPGMEGTALRMWDTRSPGLLGTPLD-KTQNVHVELEQRWHQ 641 |
| QY | 638 | ---PVEQRTFKVSOQLDWLLQLDNNTWQPHKLFSRIQ-LGLSKTYAINTLEPHQIRH 692 |
| DB | 642 | VETPLREKQVPIAPL-----HLSLEDRSERLSTHETLSLLEKELEQSQK 687 |
| QY | 693 | HKTDLLSPSGTGEVMDGVGRMSRVAKRIQVGLGDVPSVAVQGFSGAKGMWIDVD- 751 |
| DB | 688 | EASDLL-----EQNLQDQRLVAL-----GREQSAREGYVLQATC 723 |
| QY | 752 | DTGDEDWIETPSQRKWCDFDKHQRTLE-VRSVASEL--KSAGNLQLLPVLEDRARD 808 |
| DB | 724 | ERGFAWEETH--OKKIE-DLQHQHQRELEKREKDRLLAEETATATISATEAMKNAHRE 780 |
| QY | 809 | KVMRQAIGDRL-----INDLQRFSEQKHALNRPVEFQWYVESYSSRATRVSHGR 860 |
| DB | 781 | EMERELEKRSQSISSINDIEALRQYLELEQVORELEV---LSEQYSQKCLENAH-- 835 |
| QY | 861 | VPLAGLPDSOEBTLNFM--NSGDPKQKQYLQIDAWLQKRCDDTL-----KSKL 910 |
| DB | 836 | ---LAQALFAERQALQCORENQELNAHQNLNRLAEITRL--TLTGDGGGESTGL 890 |
| QY | 911 | NIRVGSAVTYMIATDPWGLENEVHVPSSKFRDEEESFTLLSCDVLNVAKSPAHFSD 970 |
| DB | 891 | PLTQGGKDAY-----ELEVLLRVK-----ESE 911 |
| QY | 971 | IQRVRAVFKPLHSLKADVIIFSTKGDVPLAKLGGDYDGDMAWCVDPPIVDGFVNAEM 1030 |
| DB | 912 | IQ-----YLKQETSSLDKEIQTARD-----KVYADKY-----KDIYELSLAXA 952 |
| QY | 1031 | PLBPDLRYLKDKTTFKOLMAHSGTGSAKQTT---YDMTKQGFHFALQPNFLGMCTN 1087 |
| DB | 953 | KADCDISRLKEQLKA-----ATEALGERSPEGTVSGYDIMKSKSN-----PDFL----- 997 |
| QY | 1088 | YKERLCYNNNSVSKPAIILSSLVGNLVDSQKQIVFNEASW 1129 |

| | | |
|----------|--|---------------------------------------|
| DB | 998 | KDRSC-----VTQLNIRSKS-----VIEQVSW 1022 |
| RESULT 5 | | |
| ID | PCLO HUMAN | STANDARD; PRT; 5147 AA. |
| AC | Q9Y6V0; Q43373; Q60305; Q9BVC8; Q9UIV2; Q9Y6U9; | |
| DT | 28-FEB-2003 (Rel. 41, Created) | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) | |
| DE | Piccolo protein (Aczonin) (Fragments). | |
| GN | PCLO OR ACZ OR KIA0059. | |
| OS | Homo sapiens (Human). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| OX | NCBI_TaxID=9606; | |
| RN | [1] | |
| RP | SEQUENCE OF 1-759 FROM N.A. | |
| RC | TISSUE=Brain; | |
| RX | MEDLINE=99439764; PubMed=10508862; | |
| RA | Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E., | |
| RA | Killmann M.W. | |
| RT | "Aczonin, a 550-kd putative scaffolding protein of presynaptic active | |
| RT | zones, shares homology regions with rim and bassoon and binds | |
| RT | profilin." | |
| RL | J. Cell Biol. 147:151-162(1999). | |
| RN | [2] | |
| RP | SEQUENCE OF 552-4404 FROM N.A. | |
| RA | Kraemer J., Wollam C., Wohldmann P., McGeehan B., | |
| RL | Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases. | |
| RN | [3] | |
| RP | SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2). | |
| RC | TISSUE=Brain; | |
| RX | MEDLINE=98290545; PubMed=9628581; | |
| RA | Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H., | |
| RA | Nomura N., Ohara O. | |
| RT | "Prediction of the coding sequences of unidentified human genes. IX. | |
| RT | The complete sequences of 100 new cDNA clones from brain which can | |
| RT | code for large proteins in vitro." | |
| RL | DNA Res. 5:31-39(1998). | |
| RN | [4] | |
| RP | SEQUENCE OF 4405-4439 FROM N.A. | |
| RC | TISSUE=Placenta; | |
| RX | MEDLINE=22388257; PubMed=12477932; | |
| RA | Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., | |
| RA | Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D., | |
| RA | Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., | |
| RA | Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hong L., | |
| RA | Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., | |
| RA | Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., | |
| RA | Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., | |
| RA | Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., | |
| RA | Bohak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., | |
| RA | Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., | |
| RA | Vallalón D.X., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., | |
| RA | Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., | |
| RA | Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., | |
| RA | Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C., | |
| RA | Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., | |
| RA | Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E., | |
| RA | Schneerch A., Schein J.E., Jones S.J.M., Marra M.A., | |
| RT | "Generation and initial analysis of more than 15,000 full-length | |
| RT | human and mouse cDNA sequences." | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). | |
| RN | [5] | |
| RP | SEQUENCE OF 4405-5147 FROM N.A. | |
| RA | Kalicki J., Elliott G. | |
| RL | Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases. | |
| CC | !- FUNCTION: May act as a scaffolding protein involved in the | |
| CC | organization of synaptic active zones and in synaptic vesicle | |
| CC | trafficking (By similarity). | |
| CC | !- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By | |

```

CC similarity).
CC -! SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9Y6V0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC VSP_003926, VSP_003927;
CC Note=No experimental confirmation available;
CC -! DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -! SIMILARITY: Contains 2 C2 domains.
CC -! SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Y19188; CAB60727.1; -
CC EMBL; AC004903; RAD20936.1; -
CC EMBL; AC004886; RAD21789.1; -
CC EMBL; AB011131; EAD35485.1; -
CC EMBL; BC001304; AAH01304.1; -
CC EMBL; AC004082; AAB97937.1; -
CC F1R; T00634; T00634.
CC HSSP; P04410; 1A25.
CC Genew; HGNC:13406; PCLO.
CC MIM; 604918; -
CC GO; GO:0005855; C:cytoskeleton; NAS.
CC GO; GO:0045202; C:synaptic junction; ISS.
CC GO; GO:0005509; F:calcium ion binding; ISS.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
CC GO; GO:0005522; F:profilin binding; ISS.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC GO; GO:0006080; P:synaptic vesicle targeting; ISS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001565; Synaptotagmin.
CC PRINTS; PR00360; C2DOMAIN.
CC PRINTS; PR00399; SYNAPTOTAGM.
CC SMART; SM00239; C2; 2.
CC PROSITE; PS00499; C2 DOMAIN 1; 1.
CC PROSITE; PS00304; C2 DOMAIN 2; 2.
CC Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
CC Repeat; Alternative splicing.
CC NON_TER 1
CC DOMAIN 400 465
CC
CC 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
CC P-A-K-P-Q-P-Q-P-X.
CC C4-TYPE (POTENTIAL).
CC C4-TYPE (POTENTIAL).
CC
CC POLY-PRO.
CC PDZ.
CC C2 DOMAIN 1.
CC C2 DOMAIN 2.
CC S -> SCNGLRIRVGGKEIPCHSGEIGAVIAKILPGSRAE
CC QTGKLMG (in isoform 2).
CC /FTId=VSP_003923.
CC K -> KPTDGTGVVSHPTGSIQ (in isoform 2).
CC /FTId=VSP_003924.
CC G -> GQMVVONAS (in isoform 2).
CC /FTId=VSP_003925.
CC TARKS -> SKRKK (in isoform 2).
CC /FTId=VSP_003926.
CC Missing (in isoform 2).
CC /FTId=VSP_003927.

```

```

SQ SEQUENCE 5147 AA; 569537 MW; CD5D84990498CD3C CRC64;
Query Match 2.2%; Score 160.5; DB 1; Length 5147;
Best Local Similarity 18.8%; Pred. No. 1.3;
Matches 286; Conservative 195; Mismatches 539; Indels 505; Gaps 72;

QY 92 PKAHADPDILPWSKEPPRAATAGQWALQTVLVLVLRNFRMPPNPTG----- 139
DB 620 PTLHGFSAK--AKQPEADSLK-----PAPPKEFSVPSQDKAPVAD 661
QY 140 -----RTFGRTLSGSLSRPTSTNTKKDKBPANVT-----ADPKRSLTRSATG- 185
DB 662 DRPKQPKWKVPTDLVSSSSAATKEDI PSSKVQSQAEKTTPLKTD SAKPSQSPFPTGB 721
QY 186 --PPIHGAALPLKFPDPVNTGSKRPSLESENQC-----TKRAKSLSDNVAA-- 232
DB 722 KVTTFDSKAIPRPSADSKII SHPGFSESSEKQKQVDPVQKKEPKKAQTMSFPKDAKPM 781
QY 233 ---AAAPPVPIASALDKVTRAHANTRDPTATGHRADOVDSF-----DTSQGT--- 278
DB 782 PKGSPTPPGPRTAGQTVTFQ--QSPKQESRRFSLNLGSLTDAPKSQPTTPQETVIG 839
QY 279 ---SYGSSVFSACRHNQST-----TQSSFPAP-----PSQ-PREKRPVDAVFEAGH 321
DB 840 KLFGFGASIFSQASNLISTAGQPHSQSGPGAPMKQAPAPSPQPTSQGPPKST-----GQ 895
QY 322 LIESPSSKG---RTTKSHIDNQLSSSSOGETSFSTYVESFPSSGEGGAIPEPSRNGLA 377
DB 896 APPAPAKSIPVKETKAPAAEKLEPKAEQAPTVKETEKKP-----PIKSKSLT 947
QY 378 RSEESARSOVQVHAPVAARLNINPKPKLWHEAPLAVAVEVTRLFMCKKVDLEDESLG 437
DB 948 AEPQKA-----VLPTKLE---KSPKESTCPL-----CKTEL---NIG 979
QY 438 LKYDPSWSTARDVTDIWKTLVRLDAPRGKPFPEKPPNDVFTAMTGNFESKSAVLSAV 497
DB 980 SKDPNPNFTC---TECKNQVCLCGF--NPTPHLTENCOTQRAISQ-----LGDI 1025
QY 498 LDVNPDSNTAPLYLVKLKPLMFQECGLTRRRGPRRPEILIPSTSTS---PSVPVTV 554
DB 1026 RKNPP--APSGP---KASPM-----PVPTESSQKTAVPQV 1057
QY 555 S-----KQGAVEEVQWLTMGQHSVLGVQWRFAFFKADG----- 589
DB 1058 KLVKQOEQVKTAEKVLKVKETLSMEKIPMWTTDQ-----KQESKLEKDSASALQ 1112
QY 590 YRKPL-----REFQLRAEDPKPIKERVHFFAETGI-----DVPKTRS 631
DB 1113 EKXPLPEEKKLPEEEKIRSEKKLLEKXKTPEDKLLPEAKTSAPQEKHDLKXQ- 1171
QY 632 VVPAEPPVEQRTFVKVSQLDWLLQLDNNNTWPHLK---LFSRIQLGLSK-----TYAI- 682
DB 1172 VQIAEKLGRVAPKTVQ-----EGKQPQTKMEGLPSGTPQSLPKEDDKTKTK 1221
QY 683 -----MTLEPHQIRH-HKTDL-----LSPSGTGEV--MNDGVGRMSRSVAKIRDLVG 727
DB 1222 EQQPPTAKPQOEKEDDKSDTSSSQPKSPGLSDTGVSSDGISS-----S 1268
QY 728 LGDVPSAVQGRFGSAKGMWVIDVDDTGDGDMWTETYPQKRWKCEDFVDKQRTLEVRVAS 787
DB 1269 LGEIPELI-----PTDEKDLKGL-----XKDSFQSSSPSS 1300
QY 788 ELKSLAGNLQLLVLEDRADKV--KMQAIGDRLINDLQRFQSKHALNPVEF---R 842
DB 1301 PSDLAKLESTVLISLEAQAQASTLADEKSEKKTQPHVSPQKQKQKTSLSLTETITSE 1360
QY 843 QWVYESYSRRATRVSHGRVFFLAGLPDSQEEFTNFMNSGDFPKKQLQDIA-----WD 897
DB 1361 ERIKESQEEKDTFK-----KDSQD---IPSSKDHKESFVDDITTRREPVD 1406
QY 898 LQKRKCDTLKSLNTRVGRSAVIYMIADFWGLNEVHVGFSS-----KFRD-----BEE 948
DB 1407 SVESESESNPVPQKGRKTS-----VGSSSSDEYKQEDSQSGSGEE 1448

```

QY 949 SFTLLSDCDVLVARSPAPHPSPDIQVRVAFKPELHSLKDVIFSTKGVDPVLAKLGGDY 1008
 Db 1449 DF-IRKQIEMGADGASSEDDEIRNQLKEISSSTESQKKEETKG---KGKITAGKH 1503
 QY 1009 -----DQDM-----AWCVNPEIVDGVNEMPLEDPLSRYLKDKTKTKQLMAS 1053
 Db 1504 RHLTRKSTSIDAGRRHSMWDEDEAFDE-----SPEL-KYRETKSQESELVVT 1554
 QY 1054 HGTGSAAKEQTTMDYMIQKSFHEALQPNFLGMCNTYKERLCVNNVSNNKPAIILSSLVGN 1113
 Db 1555 GGGGLR-----RFTIELNSTIADKYSAESS----- 1580
 QY 1114 LVDQSKQGVFNWASWAQRRELLGALSLLP-PMYKSDSMUGRPEPTIIDLKFS--- 1169
 Db 1581 ---QRKTSIYFDEE-PELEMS-----SLTDSPEDRS-----RGGSSSLHASSFTPGT 1624
 QY 1170 --IARPAIDKELEAFHNAKA-AKXTEGAGHFWDPDLASYTFFKEISDKSRSSALLFTT 1226
 Db 1625 SPTSVSILDESDSPSHKSGSKQQRKARHPHGPFLPTTDSSEBELREBELLEKEQ 1684
 QY 1227 LKNRIGVEVEKGRVLVKNKEMRDS-----KDPYP-----VR 1257
 Db 1685 EKQR--EIQOQKSSKSKKDKDLRAQRRRRPKTPPSNLSPIEDASPTSELQAAE 1742
 QY 1258 VQVYEKWCALITPEAMDKSGANYDSKVIRLLELSFLADREMTWALLLEASTAFKLYYHKS 1317
 Db 1743 MELHRSSCSYSPSIESDPGEFISPEKILIEVQ-----KVKLPFAVSLYSPDT 1792
 QY 1318 PKFVQWQAGRLAYIAKAQWTSRPGEGAPALMTAFMYAGLMDFKFTKQYVARLEGD--- 1373
 Db 1793 EQSIMGKESQKALKSAEE-----MYEEMWHTKHYKAPPAANERDEVFE 1837
 QY 1374 -----GSEYFDPDEVEVLGDDDFDG 1393
 Db 1838 KEPLYGGMLIEDYVIESLVEDTYNG 1862

RESULT 6

KMLS BOVIN
 ID KMLS_BOVIN STANDARD; PRT; 1176 AA.
 AC Q28824;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)
 DE [Contains: Telokin].
 GN MYLK.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Stomach;
 RX MEDLINE=93203148; PubMed=1284247;
 RA Kobayashi H., Inoue A., Mikawa T., Kuwayama H., Hotta Y., Masaki T.,
 Ebashi S.;
 RT "Isolation of cDNA for bovine stomach 155 kDa protein exhibiting
 RT myosin light chain kinase activity.";
 RL J. Biochem. 112:786-791(1992).
 CC -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
 CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
 CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
 CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS.
 CC -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
 CC light-chain] phosphate.
 CC -!- SUBUNIT: ISOFORM TELOKIN BINDS CALMODULIN.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=alternative initiation;
 CC Comment=At least 3 isoforms, Non-muscle, Smooth-muscle (shown
 CC here) and Telokin, are produced by alternative initiation.

CC Isoform Non-muscle is the longest and telokin is a C-terminal
 CC section without catalytic activity;
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; S57131; AAB25794.1; -;
 CC PIR; JN0583; JN0583.
 CC HSP; P56276; ITLK.
 CC InterPro; IPR008957; FN_III-like.
 CC InterPro; IPR003961; FN_III.
 CC InterPro; IPR003962; FN_III subd.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003598; Ig_c2.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR008271; Ser_thr_pkin_AS.
 CC InterPro; IPR002290; Ser_thr_pkinase.
 CC Pfam; PF00041; fn3; 1.
 CC Pfam; PF00047; ig; 2.
 CC Pfam; PF00069; pkinase; 1.
 CC PRINTS; PD00014; FNTYPEIII.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC SMART; SM00060; FN3; 1.
 CC SMART; SM00408; IGC2; 2.
 CC SMART; SM00220; S_TKC; 1.
 CC PROSITE; PS00835; IG LIKE; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC KW Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
 CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
 CC Alternative initiation.
 CC CHAIN 1 1176
 FT MYOSIN LIGHT CHAIN KINASE, ISOFORM
 FT SMOOTH-MUSCLE.
 FT MYOSIN LIGHT CHAIN KINASE, ISOFORM
 FT TELOKIN.
 FT INIT MET 1022 1022
 FT DOMAIN 100 291
 FT REPEAT 100 111
 FT REPEAT 112 123
 FT REPEAT 124 135
 FT REPEAT 136 147
 FT REPEAT 148 159
 FT REPEAT 160 171
 FT REPEAT 172 183
 FT REPEAT 184 195
 FT REPEAT 196 207
 FT REPEAT 208 219
 FT REPEAT 220 231
 FT REPEAT 232 243
 FT REPEAT 244 255
 FT REPEAT 256 267
 FT REPEAT 268 279
 FT REPEAT 280 291
 FT DOMAIN 355 444
 FT DOMAIN 498 586
 FT DOMAIN 603 673
 FT DOMAIN 725 980
 FT DOMAIN 972 1035
 FT DOMAIN 1069 1158
 FT NP BIND 731 739
 FT BINDING 754 754
 FT ACT_SITE 846 846
 FT DOMAIN 1171 1176
 FT SEQUENCE 1176 AA; 128824 MW; F53DC6D4D42D4B97 CRC64;

| | | |
|---|------|--|
| Query Match | | 2.2%; Score 160; DB 1; Length 1176; |
| Best Local Similarity | | 16.1%; Pred. No. 0.14; |
| Matches 250; Conservative 174; Mismatches 470; Indels 486; Gaps 67; | | |
| QY | 134 | PNNTP---GRTGRTLSGSPGSRPTST-----NTXKDEP---ANVTADPPK-RSLT 180 |
| Db | 121 | PAETPKPLGNVKAETPKLSTKPAETPKLSTKPAETPKLGNVKAETPKLGNK 180 |
| QY | 181 | RSATGPIHCAAIPLKPPPPVNTGSKRPLEBENLQCTYKRAKGLSDNVAAAAP---- 236 |
| Db | 181 | PTETPKPL-GSTKPAETPKPL--GSTKPAETPKPLGNVKAETPKLGNVKAETPKPLG 237 |
| QY | 237 | -----PVPIASALDKVTRRHRANTRDPTATCHRRADQVDSFDSQGTSSGVSFSACR 289 |
| Db | 238 | NVKAETPKPVNSA-----KPAETLKVQNA----- 263 |
| QY | 290 | HNQSTQSSFEAPPSOPREKRPVDAVFEAGHLIISPSKGRTRTKSHDN-----QPLS 342 |
| Db | 264 | -----KPAETPKLSNVKPA--ETPKLVGNKPAETSKP-LDNAKPAEAPKPLG 309 |
| QY | 343 | SSSQGTSTSTVYESPPSGGGAIPERSRNSGLARSSESARSQVQVHAPVVA-----RL 398 |
| Db | 310 | NAKPAE-----IPKYGKELKEIKNVCKKGAGATDSKRPES 351 |
| QY | 399 | RNIWPKPKWLHEAPLAWAVEVTRLFMHCKVDLEDESLGKYDPPSWSTARDVTDIWKTL 458 |
| Db | 352 | RGTAPTFEELQDLHVA--EGOKLLQCRV-----SSDPPATIW---- 389 |
| QY | 459 | RLDAFEGKEFPKPPNDVFVMTGNF---ESKGSANVLAVLDYNDPNSPTAPLY--LV 513 |
| Db | 390 | -----TLNGK-----TLTKTFIVLSQEGSLCSVIEKALPEBGR-----LYKQVA 430 |
| QY | 514 | KLKPLMFEQGLRTRFRGDFEFELIP-SPTSTSPSPVPPVSKPGAVVEIQWLTMGQ 572 |
| Db | 431 | KNSAGQAESSQVT-----VDVDPAPTSENAKAPEMKARPKSSLPPV---LGT 476 |
| QY | 573 | HLVGRQWPAFAKAGY-RKPLRFQLEAEDPKIILK-----ERVHFAB----- 617 |
| Db | 477 | ES-----DATVKKKAPATPKPAAMPQIIQFPEDQKVRAGESVELEFGKAGT 524 |
| QY | 618 | -----TGIFRDPVFRTKSWPAEPEVQRTBFKVSQMLDMLLQDNNWOPHLKLFERI 672 |
| Db | 525 | QPITCTWKFQKIQDSEHI--KVENSEGSKLTI-----RAAREHGCYTL 571 |
| QY | 673 | -----QLGSKTYAIMTLEPHQIRHHTDLSPSGTGEVMDGVRMSRVAKRIR-DVLGL 728 |
| Db | 572 | VENKLSRQAQVNLAV-----DKPD--PPAGT-----PCASDIRSSSLTL 610 |
| QY | 729 | GDVPSAVQGRFGSGAMVIDDGTGDEWDIETPSQKWECDFVDKHQRTLEVRSVASE 788 |
| Db | 611 | SWYGSYDG--GSAVQSVEIWDSDVKTWK-----LAT 643 |
| QY | 789 | LKSAGLNLQ-LPLVLEDRARDKVKMRQAIGDRINDLQRFSEQKHALNRPEFRQWYE 847 |
| Db | 644 | CRSTSFNVODLLP-----DREYKFRVAIN-----VYG 671 |
| QY | 848 | SYSSRATRVSHGRVFLAGLPDSQETLNFNMNSGFDPKQKYLODIANDLQKRCDTLK 907 |
| Db | 672 | -----L7ALGKPEE-----EPKDEVEVSD-----DDEKE 705 |
| QY | 908 | SKLINRVRSAYIYMIADFVWGLENEVHVGFSGKFRDEES-----FTLLSDCV 958 |
| Db | 706 | PEVDYRTVTNTEQKVSDFYDI-EERLGSKFGQVFLVVEKTKIWAGKFFKAYS----- 760 |
| QY | 959 | LVARSFAHFPDQIRVRAVFKPLHSLKDVIFSTKGDVPLAKKL-SGGDVGDMANVCW 1017 |
| Db | 761 | -----AKENIROIESTIMNCLHHPKLVQCVD--FEKANIVMLVLEIVSGG----- 805 |
| QY | 1018 | DPEIVDGVNAMPPI-EPDLSRYLKKDKTTFKQLMASHGTGSAKEQTTDMIQSPHFA 1076 |
| Db | 806 | -----ELFERIIDEDFELTERECIKYMKQ-----ISEGVEYHKQGVHLD 846 |

| | | |
|-----------|---|---|
| QY | 1077 | LQPNFLGMCTNYKRLCYINNSVSNKPAIILSSLVGNLVDSKOGIVFNBSMAQ---LR 1133 |
| Db | 847 | LKPENI-MCV-----NKTGTRIKLIDFGLARLENAGSLKVLFGTPEFVAPEVIN 895 |
| QY | 1134 | REILGGALSPLDPMVKSDSW-LGRGEPTHLIDYLFKSIARPAI--DKELEAFHNAKAAKO 1191 |
| Db | 896 | YEPFG-----YATDMMSIG-----VICVLVSGLSFFMGDNNEITLVNVTATWD 940 |
| QY | 1192 | TEDQAHFWDPLASYTYFFKEISDKSR--SAILFTTLKNRIGVE-----KEYGRL 1241 |
| Db | 941 | FDDEA-----FDEISDDAKDFISLLKKOMKQNLNCTQCLQHPWLMKDTKNM 987 |
| QY | 1242 | VKNEMRDSKDPYVRVQNVVQKWCATPEAMDKSG--ANYDSKVIKLLLELSEFLADREN 1299 |
| Db | 988 | EAKLSKDRMKKWAR-----RKW-----OKTGNVAVRAIGRLSSNAMISGLSRSS 1034 |
| QY | 1300 | TWALLRSTAFKLYYHKSPPKVMQAGRLAYIKAQWTSRPGEGAPALMTAFMYAGLMPD 1359 |
| Db | 1035 | TGSPSTPLNAEKLSESDVSQAFLEAVAEKPHVKP-----YFSKTIRD 1077 |
| QY | 1360 | KKFTKQVVARLEGDSYDPPEVY-----EVLGDDD 1390 |
| Db | 1078 | LEVVEGSAARPDCKIEGYPPEVWVFKDQDSIRESHRHQIDYDEDCGNCSLIISDVCGDDD 1137 |
| RESULT 7 | | |
| Z01_HUMAN | | |
| ID | Z01_HUMAN | STANDARD; PRT; 1736 AA. |
| AC | Q07157 | 1994 (Rel. 30, Created) |
| DT | 01-OCT-1994 | (Rel. 30, Last sequence update) |
| DT | 10-OCT-2003 | (Rel. 42, Last annotation update) |
| DE | Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona occludens 1 protein) (Tight junction protein 1). | |
| GN | TJP1 OR Z01 | |
| OS | Homo sapiens (Human) | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| OX | NCBI_TaxID=9606; | |
| RP | [1] | |
| RN | SEQUENCE FROM N.A. | |
| RC | TISUE=Liver; | |
| EC | MEDLINE=33361541; PubMed=8395056; | |
| RA | Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C., Anderson J.M.; | |
| RT | The tight junction protein ZO-1 is homologous to the Drosophila discs-large tumor suppressor protein of septate junctions."; | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838(1993). | |
| RN | [2] | |
| RP | INTERACTION WITH CGN. | |
| EX | MEDLINE=22140336; PubMed=12023291; | |
| RA | D'Atri F., Nadalutti P., Citi S.; | |
| RT | Evidence for a functional interaction between cingulin and ZO-1 in cultured cells."; | |
| RL | J. Biol. Chem. 277:27757-27764(2002). | |
| CC | -I- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN MIGHT BE INVOLVED IN STABILIZING JUNCTIONS. | |
| CC | -I- SUBUNIT: Interacts with occludin, claudins, TJP3/ZO-3 and cingulin. | |
| CC | -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE. MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT. | |
| CC | -I- ALTERNATIVE PRODUCTS: | |
| CC | Event=Alternative splicing; Named isoforms=2; | |
| CC | Name=Long; | |
| CC | isoId=Q07157-1; Sequence=Displayed; | |
| CC | Name=Short; | |
| CC | isoId=Q07157-2; Sequence=VSP_003148; | |
| CC | -I- TISSUE SPECIFICITY: THE ALPHA-CONTAINING ISOFORM IS FOUND IN MOST EPITHELIAL CELLS AND THE HIGHLY SPECIALIZED EPITHELIAL JUNCTIONS | |

CC OF RENAL GLOMERULI AND SERTOLI CELLS OF THE SEMINIFEROUS TUBULES.
 CC -!- PTM; Phosphorylated.
 CC -!- SIMILARITY: Belongs to the MAGUK family.
 CC -!- SIMILARITY: Contains 3 PDZ/DHR domains.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -!- SIMILARITY: Contains 1 guanylate kinase-like domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L14837; AA02891.1; -.
 CC FIR; A47747; A47747.
 CC HSP; P31016; I3FE.
 CC Genew; HGNC:11827; TUP1.
 CC MIM; 601009; -.
 CC GO; GO:0005913; C:cell-cell adherens junction; TAS.
 CC GO; GO:0005923; C:tight junction; TAS.
 CC GO; GO:0007043; P:intercellular junction assembly; TAS.
 CC InterPro; IPR008144; Guanylt/Ca.
 CC InterPro; IPR008145; Guanylt/Ca.
 CC InterPro; IPR001452; SH3.
 CC InterPro; IPR005417; ZonOcculDens.
 CC InterPro; IPR005418; ZonOcculS.
 CC InterPro; IPR000906; ZUS; 1.
 CC Pfam; PF00625; Guanylate_kin; 1.
 CC Pfam; PF00595; PDZ; 3.
 CC Pfam; PF00791; ZUS; 1.
 CC PRINTS; PRO1597; ZONOCCLUDNS.
 CC PRINTS; PRO1598; ZONOCCLUDNS1.
 CC SMART; SM00072; GUKC; 1.
 CC SMART; SM00228; PDZ; 3.
 CC SMART; SM00326; SH3; 1.
 CC SMART; SM00218; ZUS; 1.
 CC PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
 CC PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
 CC PROSITE; PS50106; PDZ; 3.
 CC PROSITE; PS50002; SH3; 1.
 CC Tight junction; SH3 domain; Alternative splicing; Repeat; Membrane;
 KW Phosphorylation.
 FT DOMAIN 11 98 PDZ 1.
 FT DOMAIN 174 252 PDZ 2.
 FT DOMAIN 409 490 PDZ 3.
 FT DOMAIN 504 572 SH3.
 FT DOMAIN 632 782 GUANYLATE_KINASE.
 FT DOMAIN 1231 1236 POLY-PRO.
 FT DOMAIN 1414 1420 POLY-PRO.
 FT VARSPLIC 910 989 Missing (in isoform Short).
 FT /FTID=VSP_003148.
 SQ SEQUENCE 1736 AA; 194721 MW; 508D01B7A0814PFE CRC64;
 Query Match 2.1%; Score 158.5; DB 1; Length 1736;
 Best Local Similarity 19.9%; Pred No. 0.32;
 Matches 138; Conservative 83; Mismatches 239; Indels 233; Gaps 36;
 QY 99 DTLPSKEPPRAATAGQWALQTVLLEVLNRMPPPN-NTGRTFGRSLGSPGLSRPTS 157
 DB 1129 DSRPRYEQPRASA-----LFHEBPAPGYDTHGRL--RPEAQHPSPAGPKP 1173
 QY 158 TTKRKDEPANTFAD-PEKSLTRSATCPPIHGAAPLKPDPVNTGSKRPSLESLN 216
 DB 1174 AESQKFEQYSYQVPPQGTSTAGHFEPLHGAA---AVPLIPSSQHP-----EALP 1226
 QY 217 QCTKRAKGLSDVAAAAPPVPIASALDKVP-----TRRHANTRDPTATGHRA 266
 DB 1227 SNTK-----PLPPEPTQTEBEDPAMKPSQVLTTRVKMFENKRGASLETQDV 1273
 QY 267 DQVDSFDTSCQTS--YGVSSVFSACHRNQSTTSSE-----APPSQREKRPVDAT 315

DB 1274 NTGSGFKPEVASKPSGAPIGP-----KPTSONQSEHDKTLRIPEKQPKLKEAPL 1327
 QY 316 VFEAGHL-----IESPSKGRITKSHIDNQPLSSSSQSGTGFSTYY 355
 DB 1328 IVRSNNHYDPEDEEYVRKQLSYFDRRSFENKPPAHIAASHLSEPAKPAHSQNSPSSY- 1386
 QY 356 ESFPSSGGGAIPBSRSNGLARS-EESARSOVOVHAPVVAARLNIWPKFKLHEAPL 414
 DB 1387 ----SSKG-----KPEADGVDRSPGKRYEPIQATPPP-----PPLFSQYAPSQ 1428
 QY 415 AVAMEVTRILMECK-----VDLE-DESLGLKYDPGNSWSTARDVDIWKTLRLDAPRG 465
 DB 1429 PVT--SASLHISKGAHGSGNSVSLDFQNSLVSKDPDPFS-----QN 1468
 QY 466 KPFPKPPN--DVFTMTNGNFESKSAVLSAVLDYNDNSP-----TAPLYLV 513
 DB 1469 KPATFPNRED--TAQAATYPOKSP-----PDKAPVNGTEQTQVTPAY-N 1513
 QY 514 KLKPLMFEQCCR-LTRRFQDFRFFILIPSTSPSPVPPVSKQFAGVVEIOMLTMQ 572
 DB 1514 RPTKPYTSSARPFERKPSKFNENLLPSETAHKPDLS--SKTPTSPKTLVK-----S 1565
 QY 573 HSLVGRQWRAFFAKDAGYKPLREQLRAEDPK-----PIIKERSVH----- 613
 DB 1566 HSL-----AQPPEFDSGVETFSIHAEKPKYQINNISTVPKAIPVSPSAVEDEDED 1616
 QY 614 -----FFAETGITFRPDVFKRSVVPAEPVEQRTFEKVSQMLDWL 654
 DB 1617 GHTVATARGIENSGVLSIETGVS-----IIIPQAI--EGVEQEIYFKVCRDNSIL 1670
 QY 655 LQLDNNWQ-----PH-LKLFSLIQLGL 676
 DB 1671 PLDKREKGETLLSPLVMCGPHGLKPLKPVLESL 1703
 RESULT 8
 TACC2_HUMAN STANDARD; PRT; 1026 AA.
 AC O95359; Q9NZ41; Q9NZR5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transforming acidic coiled-coil-containing protein 2 (Anti Zuai-1)
 DE (AZU-1).
 GN TACC2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain, Fetal brain, and Skeletal muscle;
 RX MEDLINE=20570483; PubMed=1121038;
 RA Gergely F., Karlsson C., Still I.H., Cowell J.K., Kilmartin J.,
 RA Raff J.W.;
 RT "The TACC domain identifies a family of centrosomal proteins that can
 RT interact with microtubules";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14352-14357(2000).
 RN [2]
 RP SEQUENCE OF 375-1026 FROM N.A. (ISOFORM 2).
 RX MEDLINE=20214826; PubMed=10749935;
 RA Chen H.-M., Schmeichel K.L., Mian I.S., Lelievre S., Petersen O.W.,
 RA Bissell M.J.;
 RT "AZU-1: a candidate breast tumor suppressor and biomarker for tumor
 RT progression";
 RL Mol. Biol. Cell 11:1357-1367(2000).
 RN [3]
 RP SEQUENCE OF 44-1026 FROM N.A. (ISOFORM 3).
 RA Pu J., Li C., Rodriguez M., Banerjee D.;
 RT "Expression of TACC2 protein mRNA in human microvascular endothelial
 RT cells";
 RL Submitted (WAY-2000) to the EMBL/GenBank/DBJ databases.

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INTERACTION WITH
RP PPARA; PPARA; RARA; RXRA; ESR1; ESR2 AND THR3.
RC TISSUE=Liver;
RX MEDLINE=20250907; PubMed=10788465;
RA Zhu Y.-J., Kan L., Qi C., Kanwar Y.S., Yeldandi A.V., Rao M.S.,
RA Reddy J.K.,
RT Isolation and characterization of peroxisome proliferator-activated
RT receptor (PPAR) interacting protein (PRIP) as a coactivator for
RT PPAR.
RL J. Biol. Chem. 275:13510-13516(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyx S.W.,
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettunen M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Rutterfield Y.S.N., Krzywicki M.I., Skalska U., Smalius D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 785-1142 FROM N.A. (ISOFORM 1), INTERACTION WITH PPARA;
RP PPARA; ESR1; ESR2; THRA AND THR3, AND MUTAGENESIS OF LEU-891 AND
RP LEU-894.
RC TISSUE=Embryo;
RX MEDLINE=20148724; PubMed=10681503;
RA Cairra F., Antonson P., Peltto-Huikko M., Treuter E., Gustafsson J.-A.,
RT "Cloning and characterization of RAP250, a nuclear receptor
RT coactivator."
RL J. Biol. Chem. 275:5308-5317(2000).
RN [4]
RP INTERACTION WITH RNP2C.
RX MEDLINE=21638469; PubMed=11704680;
RA Jung D.-J., Na S.-Y., Na D.S., Lee J.W.,
RT "Molecular cloning and characterization of CAPER, a novel coactivator
RT of activating protein-1 and estrogen receptors."
RL J. Biol. Chem. 277:1229-1234(2002).
RN [5]
RP -1- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
RP receptors and stimulates the transcriptional activities in a
RP hormone-dependent fashion. Coactivates expression in an agonist-
RP and AF2-dependent manner. Involved in the coactivation of
RP different nuclear receptors, such as for steroids (GR and ERs),
RP retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR)
RP and prostanooids (PPARs). Probably functions as a general
RP coactivator, rather than just a nuclear receptor coactivator. May
RP also be involved in the coactivation of the NF-kappa-B pathway.
RP May coactivate expression via a remodeling of chromatin and its
RP interaction with histone acetyltransferase proteins. Involved in
RP placental, cardiac, hepatic and embryonic development.
RP -1- SUBUNIT: Monomer and homodimer. Interacts in vitro with the basal
RP transcription factors GTF2A and TBP, suggesting an autonomous
RP transactivation function. Interacts with NCOAL, CRSP3, RRM14, the
RP histone acetyltransferase proteins NCOA6IP and ESR300 and CREBBP, and with
RP methyltransferase proteins NCOA6IP and HRM111 (By similarity).
RP Interacts with RNP2C. Belongs to the ASC-2/NCOA6 complex (ASCOM),
RP which contains ASC-2/NCOA6, the retinoblastoma-binding protein

CC RRO-3/ RBSP5, alpha- and beta-tubulins, the trithorax group
CC PROTEINS MLL2 and MLL3, and ASH2/ASCL2 (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JUL19-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JUL19-2; Sequence=VSP_003410;
CC Note=Acts as a dominant negative repressor;
CC TISSUE SPECIFICITY: Widely expressed. High expression in testis
CC and weak expression in small intestine.
CC -1- DEVELOPMENTAL STAGE: Expressed at E9 in placenta and at weaker
CC level in uterus. High expression in neural tube and in CNS
CC throughout development. High expression in sensory ganglia and
CC retina from E11. In the alimentary tract and olfactory epithelium
CC expression was seen from E13. Strong expression present in liver
CC and kidney, from E11 and E13 respectively, and then expression in
CC decreased at later stages of development. Moderate expression in
CC lung from E13, while it decreases during postnatal life. Strong
CC expression in thymus from E15 onwards, and in spleen from E17 and
CC during early postnatal life, then, the expression decreases.
CC -1- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs. Only
CC motif 1 is essential for the association with nuclear receptors.
CC -1- PTM: Phosphorylated (By similarity).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF216186; AAF35860.1; -
CC EMBL: BC031113; AAF31113.1; -
CC EMBL: AF135169; AAF35973.1; -
CC MGD: MGI:1929915; Ncoaf.
CC GO: GO:0005634; C:nucleus; IDA.
CC GO: GO:0005667; C:transcription factor complex; IDA.
CC GO: GO:0003682; F:chromatin binding; IDA.
CC GO: GO:0003031; F:estrogen receptor binding; ISS.
CC GO: GO:0046965; F:retinoid X receptor binding; ISS.
CC GO: GO:0046966; F:thyroid hormone receptor binding; ISS.
CC GO: GO:0003713; F:transcription co-activator activity; ISS.
CC GO: GO:0016563; F:transcriptional activator activity; IDA.
CC GO: GO:0007420; P:brain development; IMP.
CC GO: GO:0001701; P:embryonic development (sensu Mammalia); IMP.
CC GO: GO:0007507; P:heart development; IMP.
CC GO: GO:0030099; P:myeloid blood cell differentiation; ISS.
CC GO: GO:0006367; P:transcription initiation from Pol II promoter; IDA.
CC KW Transcription regulation; Activator; Nuclear protein; Repeat;
CC Alternative splicing.
CC DOMAIN 1 1060 CREBBP-BINDING REGION (BY SIMILARITY).
CC DOMAIN 1 932 TBF/GTF2A-BINDING REGION (BY SIMILARITY).
CC DOMAIN 1 1314 NCOA1-BINDING REGION (BY SIMILARITY).
CC DOMAIN 777 931 NCOA6IP-BINDING REGION (BY SIMILARITY).
CC DOMAIN 1644 2067 EP300/CRSP3-BINDING REGION
CC (BY SIMILARITY).
CC DOMAIN 227 1044 GLN-RICH.
CC DOMAIN 376 381 POLY-PRO.
CC DOMAIN 917 922 POLY-LYS.
CC DOMAIN 1543 1592 SER-RICH.
CC DOMAIN 891 895 LXXLL MOTIF 1.
CC SITE 1495 1499 LXXLL MOTIF 2.
CC SITE 458 2067 Missing (in isoform 2).
CC VARSPLIC 891 894 /FTId=VSP_003410.
CC MUTAGEN L1NL-AVNA: ABOLISHES INTERACTION WITH
CC NUCLEAR RECEPTORS.
CC G -> S (IN REF. 2).
CC W -> R (IN REF. 2).
CC M -> I (IN REF. 2).
CC Q -> QQ (IN REF. 2).
CC CONFLICT 39 39
CC CONFLICT 109 109
CC CONFLICT 194 194
CC CONFLICT 290 290

```
FT CONFLICT 1014 1014 P -> L (IN REF. 3).
FT CONFLICT 1141 1142 SE -> RS (IN REF. 3).
SQ SEQUENCE 2067 AA; 213663 MW; C855P8777157AD48 CRC64;

Query Match
Best Local Similarity 20.8%; Pred. No. 0.58; Length 2067;
Matches 130; Conservative 74; Mismatches 246; Indels 176; Gaps 30;

QY 3 PITPKRNSPVEEIIINRLNNDYGLQCVADTTLTPHRRKELAEDEDFGRHDKIYRALN 62
Db 1063 PLNPSQRMPEVQ--SGNVPMVWGLQGPASVPSPDQKPMNSVNTPMGNSR-----K 1114
QY 63 FLYWRKDDSLAQAEAFIEAASNSVWPKAHADPDILP-----WKEP--- 107
Db 1115 MYIQENPQNSSPSGESLSEARSGSEVPSVAGGPNMPSHLVVSQQLMWTGPKGPS 1174
QY 108 PRAATAG--QQWALQTVLLEVMFPPNNTP---GRTFORTLSGSGLSRPTSTNTK 161
Db 1175 PLSATQATPQ-----PPVNSLPSSGHGF-PNVAAATQTSRKPFPURA 1218
QY 162 R-----KDEPANVTADPPKSLRSATGPPHGAIAIPKTPDPVNTG-SKRPSE---E 211
Db 1219 SPRYPYQPPNN---RPPSTEPSISLSPEELNASIAGLPFPQINIPLPNPNLRGFD 1274
QY 212 SENLAQCTKRAKGLSDNVAAAP--PVPIASALDKVPTRRHANTRDPTATGHRADQV 269
Db 1275 QQLNPTTLKALGQAPSNLTITNPNFAAPQAHKLDVSVNSGKSNPCIT--KRASPS 1331
QY 270 DSDFTSQTSQSSVFSACRNQSTQTSFEAPPQPREKP-----VDATVFEAGHL 322
Db 1332 NRRSSPSS-----RKTTPS-----PGRQNSKAPKLTASQSTTLMQN--- 1371
QY 323 IESPSKGRITKTHIDNQLSSSSQGETSFSTVYESFSSGG---EGALPEPSRSNGLAR 378
Db 1372 MELPRNLVGPFLANPELPG-----SFPNNTGLNPNQPVFPVPAAGTVLED 1418
QY 379 SEESA-----RSQVQHPVAPVAAARLNIWPKPKMLHEAPLAVAMEVTRLF 424
Db 1419 NKESVNIPODSCQNAQGRKEQVNTLKVPTQ-----EAKMAVP----- 1458
QY 425 MHCKVDLESL--GLKYDPS--WSTARVDVTDIKTLR-----LDAFGKFPPEKP 472
Db 1459 -----EQSKKQGPDPKPLPSVEENKMLSPAMREAPTSLSQLLDNSGANVTIKP 1511
QY 473 P--NDVFVTAMTGNFESKGSAYVLVSAVLDPNDSPTAPLYLVKPLKPLMFQGGQLTRF 530
Db 1512 PGLTDLEVTPPVVSGEDLRKASVITLQD-PSKEPSTSL---SSPHSSSEPCSTLAR-- 1564
QY 531 GPDREFEILPSTSTSPVPPVWK 556
Db 1565 -----SELSEVSSNAAPSIPVMSR 1584

RESULT 10
PCLO CHICK
ID PCLO CHICK STANDARD; PRT; 5120 AA.
AC Q9PU36;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Piccolo protein (Aczonin) (Fragment).
GN PCLO OR ACZ.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
Killmann M.W.;
```

```
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin."
RL J. Cell Biol. 147:151-162(1999)
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y19187; CAB60725.1; -
CC HSPSP; P04410; 1A25.
CC GO; GO:0045202; C:synaptic junction; ISS.
CC GO; GO:0005509; F:calcium ion binding; ISS.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
CC GO; GO:0005522; F:profilin binding; ISS.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001565; Synaptotagmin.
CC InterPro; IPR008899; Znf_Piccolo.
CC Pfam; PF00168; C2; 2.
CC Pfam; PF00595; PDZ; 1.
CC PRINTS; PR00399; SYNAPTOTAGMN.
CC SMART; SM00239; C2; 2.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00499; C2_DOMAIN_1; 1.
CC PROSITE; PS00004; C2_DOMAIN_2; 2.
CC PROSITE; PS0106; PDZ; 1.
CC Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
CC Repeat.
CC NON_TER 1 1
CC DOMAIN 258 357 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
CC P-A-K-P-Q-P-Q-P-X.
CC ZN_FING 368 392 C4-TYPE (POTENTIAL).
CC ZN_FING 836 859 C4-TYPE (POTENTIAL).
CC DOMAIN 2324 2343 POLY-PRO.
CC DOMAIN 4414 4493 PDZ.
CC DOMAIN 4627 4726 C2 DOMAIN 1.
CC DOMAIN 5003 5094 C2 DOMAIN 2.
CC SEQUENCE 5120 AA; 560751 MW; A658D9891B65B412 CRC64;

Query Match
Best Local Similarity 18.7%; Pred. No. 2.9;
Matches 267; Conservative 201; Mismatches 524; Indels 435; Gaps 73;

QY 99 DTLPWSEK-----PRAATAGQWALQTVLLEVMFPP-----PNNTPGRT 141
Db 542 DTVPTSAAGVKQDLADPQSPST-QQKVTDSMPETTK---PPADTHPAGDKPDSKPLPQ 597
QY 142 FGRTLSGSPGLSR---PTSTNTKRKDEPANVTADPPKSLRSATGPPHGAIAIPKFP 198
Db 598 VSRQKSDPKLASQSGAKSDAKTKPSEPAV--KDPKQLQTKPAKPKTKPAP---KGP 652
QY 199 DPVNTG-----SKRPSLESENLCQCTKRAKGLSDNVAAAPVPTASALDKVPTRRHN 254
Db 653 Q-AGTGPRTPTSAQAPQPOQ--FKTPEQRRRSLNLGIGTDAPKP-----QPT----- 698
```


RP SEQUENCE OF 3734-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara C.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain";
RL DNA Res. 4:345-349(1997).
CC -!- FUNCTION: F-actin-binding protein which may play a role in cross-
CC linking actin to other cytoskeletal proteins. Also binds to
CC microtubules (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=2;
CC isoId=Q9UPN3-2; Sequence=Displayed;
CC Name=1;
CC isoId=Q9UPN3-1; Sequence=VSP_007341;
CC Name=3;
CC isoId=Q9UPN3-3; Sequence=Not described;
CC Name=4;
CC isoId=Q96PK2-1; Sequence=External;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC -!- SIMILARITY: Contains 1 actin-binding domain.
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 37 spectrin repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB029290; BAA83821.1; -;
CC EMBL; AF141968; AAF06360.1; -;
CC EMBL; AF325341; AAL39000.1; -;
CC EMBL; AF325330; AAL39000.1; JOINED.
CC EMBL; AF325331; AAL39000.1; JOINED.
CC EMBL; AF325332; AAL39000.1; JOINED.
CC EMBL; AF325333; AAL39000.1; JOINED.
CC EMBL; AF325334; AAL39000.1; JOINED.
CC EMBL; AF325335; AAL39000.1; JOINED.
CC EMBL; AF325336; AAL39000.1; JOINED.
CC EMBL; AF325339; AAL39000.1; JOINED.
CC EMBL; AF325340; AAL39000.1; JOINED.
CC EMBL; AB033077; BAA86565.1; -;
CC EMBL; AL137853; CAC15920.1; -;
CC EMBL; AB007934; BAA32310.2; -;
CC PIR; T00079; T00079.
CC HSSP; Q01082; 1BKR.
CC Genew; HGNC:13664; MACF1.
CC GO; GO:0005856; Cytoskeleton; NAS.
CC GO; GO:0003780; Fractin cross-linking activity; NAS.
CC GO; GO:0005089; Fractin ion binding; NAS.
CC GO; GO:0008017; F-actin binding; NAS.
CC InterPro; IPR001589; Actinbind_actin.
CC InterPro; IPR001715; Calponin-like.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR003108; GAS2.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00036; ehand; 2.
CC Pfam; PF02187; GAS2; 1.
CC Pfam; PF00435; spectrin; 27.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00033; CH; 2.
CC SMART; SM00054; Efh; 2.

DR SMART; SM00243; GAS2; 1.
DR SMART; SM00150; SPEC; 36.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00002; SH3; FALSE_NEG.
KW Actin-binding; Cytoskeleton; Calcium; Calcium-binding; Repeat;
KW SH3 domain; Coiled coil; Alternative splicing.
FT DOMAIN 1 295 ACTIN-BINDING.
FT DOMAIN 78 181 CH 1.
FT DOMAIN 194 295 CH 2.
FT DOMAIN 243 265 COILED COIL (POTENTIAL).
FT DOMAIN 477 529 COILED COIL (POTENTIAL).
FT DOMAIN 722 751 COILED COIL (POTENTIAL).
FT DOMAIN 816 843 COILED COIL (POTENTIAL).
FT DOMAIN 1013 1118 COILED COIL (POTENTIAL).
FT DOMAIN 1164 1191 COILED COIL (POTENTIAL).
FT DOMAIN 1399 1690 COILED COIL (POTENTIAL).
FT DOMAIN 1780 1843 COILED COIL (POTENTIAL).
FT DOMAIN 1975 2005 COILED COIL (POTENTIAL).
FT DOMAIN 2039 2312 COILED COIL (POTENTIAL).
FT DOMAIN 2385 2417 COILED COIL (POTENTIAL).
FT DOMAIN 2544 2695 COILED COIL (POTENTIAL).
FT DOMAIN 2760 2838 COILED COIL (POTENTIAL).
FT DOMAIN 2911 3001 COILED COIL (POTENTIAL).
FT DOMAIN 3130 3164 COILED COIL (POTENTIAL).
FT DOMAIN 3244 3277 COILED COIL (POTENTIAL).
FT DOMAIN 3418 3482 COILED COIL (POTENTIAL).
FT DOMAIN 3596 3666 COILED COIL (POTENTIAL).
FT DOMAIN 3786 3806 COILED COIL (POTENTIAL).
FT DOMAIN 3852 3931 COILED COIL (POTENTIAL).
FT DOMAIN 3967 3987 COILED COIL (POTENTIAL).
FT DOMAIN 4084 4218 COILED COIL (POTENTIAL).
FT DOMAIN 4343 4378 COILED COIL (POTENTIAL).
FT DOMAIN 4408 4437 COILED COIL (POTENTIAL).
FT DOMAIN 4468 4498 COILED COIL (POTENTIAL).
FT DOMAIN 4907 4935 COILED COIL (POTENTIAL).
FT DOMAIN 5044 5067 COILED COIL (POTENTIAL).
FT REPEAT 314 355 SPECTRIN 1.
FT REPEAT 591 623 SPECTRIN 2.
FT REPEAT 680 784 SPECTRIN 3.
FT REPEAT 786 800 SPECTRIN 4.
FT REPEAT 871 923 SH3
FT REPEAT 1250 1272 SPECTRIN 5.
FT REPEAT 1287 1342 SPECTRIN 6.
FT REPEAT 1455 1534 SPECTRIN 7.
FT REPEAT 1547 1659 SPECTRIN 8.
FT REPEAT 1815 1891 SPECTRIN 9.
FT REPEAT 1932 2042 SPECTRIN 10.
FT REPEAT 2260 2280 SPECTRIN 11.
FT REPEAT 2372 2395 SPECTRIN 12.
FT REPEAT 2398 2507 SPECTRIN 13.
FT REPEAT 2510 2618 SPECTRIN 14.
FT REPEAT 2621 2728 SPECTRIN 15.
FT REPEAT 2731 2838 SPECTRIN 16.
FT REPEAT 2841 2945 SPECTRIN 17.
FT REPEAT 2987 3024 SPECTRIN 18.
FT REPEAT 3136 3163 SPECTRIN 19.
FT REPEAT 3187 3274 SPECTRIN 20.
FT REPEAT 3277 3383 SPECTRIN 21.
FT REPEAT 3386 3492 SPECTRIN 22.
FT REPEAT 3495 3601 SPECTRIN 23.
FT REPEAT 3604 3673 SPECTRIN 24.
FT REPEAT 3713 3819 SPECTRIN 25.
FT REPEAT 3832 3927 SPECTRIN 26.
FT REPEAT 3982 4043 SPECTRIN 27.
FT REPEAT 4046 4152 SPECTRIN 28.
FT REPEAT 4155 4262 SPECTRIN 29.

Query Match 2.1%; Score 153; DB 1; Length 5430;
Best Local Similarity 17.5%; Pred. No. 3.9;
Matches 217; Conservative 175; Mismatches 415; Indels 436; Gaps 58;

QY 264 RADVDSDTSQGTSSVGSFVACRHNQSTQSSFPAPSPQPREKPPVATVFEAGHLI 323
 DB 3765 RHDSMDL-----FHSRSEIFGTCGEBQKTVL-----QEKTESLIQOYEALSL 3809
 QY 324 ESPSGRRTTKGHI-DNQPLSSSQGTSFSTVYVESFP-----SSCGEAIPEPSRNSGLAR 378
 DB 3810 NSERVARLERAQVLNQ-----FWETYEELSPWIEBETRALIAQLPSAIDHBQLR 3859
 QY 379 SEE-----SARGQVQHAP-----VVAARLNINWPKPKWLHEAPLAVANEVTRLFHCK 428
 DB 3860 QOQEEMRQLRESIAHKEHIDKLKIGPQLKELNPEEGEMVEEK-----YQKAENMYAQIK 3915
 QY 429 VDLEDESGLKYDPSWSTARDVTDITWKTLYRLDAFRGPKPPPEKPPNDVFVTAMTGFESK 488
 DB 3916 FEVRQALAL--DEAVSQSTQITE-----FHDK-----IEPMLTLENL 3952
 QY 489 GSAYLSAVLYNDP-----NSTAPLYLVKPLMFEGOCR-----LTRRPGPR-- 534
 DB 3953 SRRLMPPLIPAEVDKIREC-SDNKSATVELEKLP-SFEALKRGRBELIGRSQGADKDL 4011
 QY 535 -----PFE-----ILPGTST----- 546
 DB 4012 AAKETQDKLDQWFFWEDIKARABEREIKFLDVLAEKFWYDVAALTIKTQDIDVHD 4071
 QY 547 --SPSPVPVSKQPGAVEVIOQLTMGQHS-----LVGRQWRAPFA-----KDAYKPKPL 594
 DB 4072 LESPGIDPSITKQVEAAETIKETDGLHEEFIRILGAD--LIFACGETEKPEVRKSI 4129
 QY 595 REFQRAEDPKPIIKERYHFAETGITRDPVFKTRSVVPAEPEVQRTQTEK--VSOQLD 652
 DB 4130 DENNAWENLNKTKWERLE-----KLEDAWQAQVQDQLQAMFD 4169
 QY 653 WLQLDNNTPHLLKFS-----RIQLGLSKTYAINTLEPHQIR-----HHKTDL 697
 DB 4170 W--LDNTV---IKLCTMPVGTDLNTVKQDNEMKEFKVEVYQ-QQIMEKLNHQGEL 4221
 QY 698 LPSSTGEVMDGV-----GRMSRSVAKIRD-----VLGLGDVPSAVQGRFGSA 742
 DB 4222 MLKXATDETRDIIEPTELKHLWENIGEKIAHQKLEGALLALGQFOHALBELMS-- 4279
 QY 743 KGMWVIDVDD-----TGDEDMTETPSQKMECDFVDKHQRTLE--VRSVASEL--KS 791
 DB 4280 ---WLTHTELLDQAQPSISGDPKVELELKHVHLKNDVLAQATVETVKNAGNELLESS 4336
 QY 792 AGNLQLLPLEDRARDKVMQQAIGDRLNDLQRFSEQKHALNRPVEFQWVYESYSS 851
 DB 4337 AGDDASSL-----RSLEAMNQWESVLQKTEEREQQLQSTLQQAQGFHSEI-EDFL 4388
 QY 852 RATRVIS--HGRVPFLAGLPD-----SQEETLNFLMN----- 880
 DB 4389 ELTRVESQLSASKP-TGGLPETAREQLDTHMELYSQKAKEETYNQLLDKRLMLLSRDD 4447
 QY 881 --SGFPDPKQKYLQDIANDLQKRCDDTLKSKUNIRVRSAYIMTADPWGLENEHVAG 938
 DB 4448 SGSGSKTEQSVALLEGKQHWVSSKMEERKSL-----EEALN 4484
 QY 939 FSKSPREDEESF---TLSDCDVLVARSAPAHFSPDIQ---RVRAVFKPELHSLKDVII-- 990
 DB 4485 LATEFQNSIQEINLWTLAQESLNTASPPSLINVLQSQIEHKVFANEVNAHRDQIEL 4544
 QY 991 -----FSTKGDVPLAKKL-----SGGDYDG-----DMAWVC 1016
 DB 4545 DQTGNQLKELSKQDVLVILKNLLSVQSRWEKVVQSR--ERGRSLDDARKRAKQFHEAW-- 4602
 QY 1017 WDPEIIVGDFVNAEMPLEPDLRYLKDKTKTTFKQLMASH-----GTGS 1058
 DB 4603 --KKLIDWLDEASHLDSLE--ISNDPKIKLQSKKEFKQKTLGGQKQPYDITIRGR 4658
 QY 1059 AAKQETTYDMIQSFALQPNFLGMCNTYKBERLYINNSVKNKPAIILSLVGNLVDQS 1118
 DB 4659 ALKEKTLPPDSQKL-----DNFLGEVRDKWDTVC-----GKSVRQ 4695

QY 1119 ---KQIIVNEASWAQLRRELGGALSLPDPWYKSDSWLGRGEPHTHIDYLFKFIARPAI 1175
 DB 4696 HKLEBALFS-----GQFMDALCALVDWLYKVEPQLAEDQPVH----- 4733
 QY 1176 DKELBAFNHAKAAKDEGAHFWDFDLASYYTFFPKETSDKSRSSALLFTTLKRNIGBE 1235
 DB 4734 -GDLDLVNMLDAHK-----VFOKE-----LGRKGTGVQ 4761
 QY 1236 --KEYGR-LVKNKEVRDSDKPPVVRVNOVYKWCACAITPEAMDK 1275
 DB 4762 VLKRSGRELIENS--RDDTTWVKQLQELSTENDTVCKLSYSK 4802
 RESULT 12
 MAC4_HUMAN
 ID MAC4_HUMAN STANDARD; PRT: 5938 AA.
 AC Q96PK2; O8WKYL;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Microtubule-actin crosslinking factor 1, isoform 4.
 GN MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX MEDLINE=21833812; PubMed=11845288;
 RA Gong T.-W.L., Besirli C.G., Lomax M.I.;
 RT "MACF1 gene structure: a hybrid of plectin and dystrophin.";
 RL Mamm. Genome 12:852-861(2001).
 CC -!- FUNCTION: May play a role in cross-linking cytoskeletal proteins
 CC by binding intermediate filaments to the N-terminal plectin
 CC repeats and microtubules to the C-terminus.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=Q96PK2-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q9UPN3-1; Sequence=External;
 CC Name=2;
 CC IsoId=Q9UPN3-2; Sequence=External;
 CC Name=3;
 CC IsoId=Q9UPN3-3; Sequence=External;
 CC -!- TISSUE SPECIFICITY: Expressed in heart, lung, pituitary and
 CC placenta, not found in brain, kidney, liver, pancreas or skeletal
 CC muscle.
 CC -!- SIMILARITY: Belongs to the plectin or cytolinker family.
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -!- SIMILARITY: Contains 19 plectrin repeats.
 CC -!- SIMILARITY: Contains 32 spectrin repeats.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AF317696; AAL09459.1; -
 CC EMBL; AF325341; AAL38997.1; -
 CC EMBL; AF325343; AAL38997.1; JOINED.
 CC EMBL; AF325334; AAL38997.1; JOINED.
 CC EMBL; AF325335; AAL38997.1; JOINED.
 CC EMBL; AF325336; AAL38997.1; JOINED.
 CC EMBL; AF325339; AAL38997.1; JOINED.
 CC EMBL; AF325340; AAL38997.1; JOINED.
 CC GO; GO:0005856; C:cytoskeleton; ISS.
 CC GO; GO:0005509; F:calcium ion binding; ISS.
 CC GO; GO:0008017; F:microtubule binding; ISS.

| | | | | | | |
|----|--|----|--------|-------|-------|---------------|
| DR | InterPro; IPR002048; EF-hand. | FT | REPEAT | 3453 | 3555 | SPECTRIN 14. |
| DR | InterPro; IPR003108; GAS2. | FT | REPEAT | 3562 | 3664 | SPECTRIN 15. |
| DR | InterPro; IPR001101; Plectrin.repeat. | FT | REPEAT | 3671 | 3775 | SPECTRIN 16. |
| DR | InterPro; IPR003017; Spectrin. | FT | REPEAT | 3782 | 3884 | SPECTRIN 17. |
| DR | Pfam; PF00036; sfhand; 2. | FT | REPEAT | 3891 | 3993 | SPECTRIN 18. |
| DR | Pfam; PF02187; GAS2; 1. | FT | REPEAT | 4000 | 4102 | SPECTRIN 19. |
| DR | Pfam; PF00681; Plectrin; 11. | FT | REPEAT | 4109 | 4211 | SPECTRIN 20. |
| DR | Pfam; PF00435; spectrin; 26. | FT | REPEAT | 4218 | 4320 | SPECTRIN 21. |
| DR | ProDom; PD000012; EF-hand; 1. | FT | REPEAT | 4327 | 4428 | SPECTRIN 22. |
| DR | SMART; SM00054; EFh; 2. | FT | REPEAT | 4438 | 4544 | SPECTRIN 23. |
| DR | SMART; SM00243; GAS2; 1. | FT | REPEAT | 4551 | 4653 | SPECTRIN 24. |
| DR | SMART; SM00250; PLEC; 19. | FT | REPEAT | 4660 | 4763 | SPECTRIN 25. |
| DR | SMART; SM00150; SPEC; 32. | FT | REPEAT | 4770 | 4872 | SPECTRIN 26. |
| DR | PROSITE; PS00018; EF HAND; 2. | FT | REPEAT | 4879 | 4982 | SPECTRIN 27. |
| DR | PROSITE; PS00018; EF HAND; 2. | FT | REPEAT | 4989 | 5091 | SPECTRIN 28. |
| KW | Cytoskeleton; Calcium; Calcium-binding; Repeat; Coiled coil; | FT | REPEAT | 5098 | 5201 | SPECTRIN 29. |
| KW | Alternative splicing. | FT | REPEAT | 5208 | 5309 | SPECTRIN 30. |
| FT | DOMAIN 1830 1936 | FT | REPEAT | 5316 | 5418 | SPECTRIN 31. |
| FT | DOMAIN 2001 2192 | FT | REPEAT | 5425 | 5555 | SPECTRIN 32. |
| FT | DOMAIN 2282 2345 | FT | REPEAT | 5555 | 5657 | SPECTRIN 33. |
| FT | DOMAIN 2477 2507 | FT | REPEAT | 5657 | 5759 | SPECTRIN 34. |
| FT | DOMAIN 2541 2654 | FT | REPEAT | 5759 | 5861 | SPECTRIN 35. |
| FT | DOMAIN 2686 2814 | FT | REPEAT | 5861 | 5963 | SPECTRIN 36. |
| FT | DOMAIN 2887 2919 | FT | REPEAT | 5963 | 6065 | SPECTRIN 37. |
| FT | DOMAIN 3046 3197 | FT | REPEAT | 6065 | 6167 | SPECTRIN 38. |
| FT | DOMAIN 3262 3303 | FT | REPEAT | 6167 | 6269 | SPECTRIN 39. |
| FT | DOMAIN 3632 3666 | FT | REPEAT | 6269 | 6371 | SPECTRIN 40. |
| FT | DOMAIN 3746 3779 | FT | REPEAT | 6371 | 6473 | SPECTRIN 41. |
| FT | DOMAIN 3920 3984 | FT | REPEAT | 6473 | 6575 | SPECTRIN 42. |
| FT | DOMAIN 4098 4168 | FT | REPEAT | 6575 | 6677 | SPECTRIN 43. |
| FT | DOMAIN 4288 4308 | FT | REPEAT | 6677 | 6779 | SPECTRIN 44. |
| FT | DOMAIN 4354 4386 | FT | REPEAT | 6779 | 6881 | SPECTRIN 45. |
| FT | DOMAIN 4397 4433 | FT | REPEAT | 6881 | 6983 | SPECTRIN 46. |
| FT | DOMAIN 4489 4589 | FT | REPEAT | 6983 | 7085 | SPECTRIN 47. |
| FT | DOMAIN 4586 4720 | FT | REPEAT | 7085 | 7187 | SPECTRIN 48. |
| FT | DOMAIN 4845 4880 | FT | REPEAT | 7187 | 7289 | SPECTRIN 49. |
| FT | DOMAIN 4910 4939 | FT | REPEAT | 7289 | 7391 | SPECTRIN 50. |
| FT | DOMAIN 4970 5000 | FT | REPEAT | 7391 | 7493 | SPECTRIN 51. |
| FT | DOMAIN 5409 5437 | FT | REPEAT | 7493 | 7595 | SPECTRIN 52. |
| FT | DOMAIN 5546 5569 | FT | REPEAT | 7595 | 7697 | SPECTRIN 53. |
| FT | DOMAIN 5598 5610 | FT | REPEAT | 7697 | 7799 | SPECTRIN 54. |
| FT | CA_BIND 5634 5646 | FT | REPEAT | 7799 | 7901 | SPECTRIN 55. |
| FT | REPEAT 12 49 | FT | REPEAT | 7901 | 8003 | SPECTRIN 56. |
| FT | REPEAT 53 88 | FT | REPEAT | 8003 | 8105 | SPECTRIN 57. |
| FT | REPEAT 89 126 | FT | REPEAT | 8105 | 8207 | SPECTRIN 58. |
| FT | REPEAT 130 164 | FT | REPEAT | 8207 | 8309 | SPECTRIN 59. |
| FT | REPEAT 166 202 | FT | REPEAT | 8309 | 8411 | SPECTRIN 60. |
| FT | REPEAT 203 240 | FT | REPEAT | 8411 | 8513 | SPECTRIN 61. |
| FT | REPEAT 243 278 | FT | REPEAT | 8513 | 8615 | SPECTRIN 62. |
| FT | REPEAT 279 316 | FT | REPEAT | 8615 | 8717 | SPECTRIN 63. |
| FT | REPEAT 318 354 | FT | REPEAT | 8717 | 8819 | SPECTRIN 64. |
| FT | REPEAT 375 411 | FT | REPEAT | 8819 | 8921 | SPECTRIN 65. |
| FT | REPEAT 438 474 | FT | REPEAT | 8921 | 9023 | SPECTRIN 66. |
| FT | REPEAT 497 533 | FT | REPEAT | 9023 | 9125 | SPECTRIN 67. |
| FT | REPEAT 556 592 | FT | REPEAT | 9125 | 9227 | SPECTRIN 68. |
| FT | REPEAT 615 651 | FT | REPEAT | 9227 | 9329 | SPECTRIN 69. |
| FT | REPEAT 674 710 | FT | REPEAT | 9329 | 9431 | SPECTRIN 70. |
| FT | REPEAT 733 769 | FT | REPEAT | 9431 | 9533 | SPECTRIN 71. |
| FT | REPEAT 792 828 | FT | REPEAT | 9533 | 9635 | SPECTRIN 72. |
| FT | REPEAT 851 887 | FT | REPEAT | 9635 | 9737 | SPECTRIN 73. |
| FT | REPEAT 910 946 | FT | REPEAT | 9737 | 9839 | SPECTRIN 74. |
| FT | REPEAT 969 1005 | FT | REPEAT | 9839 | 9941 | SPECTRIN 75. |
| FT | REPEAT 1028 1064 | FT | REPEAT | 9941 | 10043 | SPECTRIN 76. |
| FT | REPEAT 1087 1123 | FT | REPEAT | 10043 | 10145 | SPECTRIN 77. |
| FT | REPEAT 1146 1182 | FT | REPEAT | 10145 | 10247 | SPECTRIN 78. |
| FT | REPEAT 1205 1241 | FT | REPEAT | 10247 | 10349 | SPECTRIN 79. |
| FT | REPEAT 1264 1300 | FT | REPEAT | 10349 | 10451 | SPECTRIN 80. |
| FT | REPEAT 1323 1359 | FT | REPEAT | 10451 | 10553 | SPECTRIN 81. |
| FT | REPEAT 1382 1418 | FT | REPEAT | 10553 | 10655 | SPECTRIN 82. |
| FT | REPEAT 1441 1477 | FT | REPEAT | 10655 | 10757 | SPECTRIN 83. |
| FT | REPEAT 1499 1535 | FT | REPEAT | 10757 | 10859 | SPECTRIN 84. |
| FT | REPEAT 1558 1594 | FT | REPEAT | 10859 | 10961 | SPECTRIN 85. |
| FT | REPEAT 1617 1653 | FT | REPEAT | 10961 | 11063 | SPECTRIN 86. |
| FT | REPEAT 1676 1712 | FT | REPEAT | 11063 | 11165 | SPECTRIN 87. |
| FT | REPEAT 1735 1771 | FT | REPEAT | 11165 | 11267 | SPECTRIN 88. |
| FT | REPEAT 1794 1830 | FT | REPEAT | 11267 | 11369 | SPECTRIN 89. |
| FT | REPEAT 1853 1889 | FT | REPEAT | 11369 | 11471 | SPECTRIN 90. |
| FT | REPEAT 1912 1948 | FT | REPEAT | 11471 | 11573 | SPECTRIN 91. |
| FT | REPEAT 1971 2007 | FT | REPEAT | 11573 | 11675 | SPECTRIN 92. |
| FT | REPEAT 2030 2066 | FT | REPEAT | 11675 | 11777 | SPECTRIN 93. |
| FT | REPEAT 2089 2125 | FT | REPEAT | 11777 | 11879 | SPECTRIN 94. |
| FT | REPEAT 2148 2184 | FT | REPEAT | 11879 | 11981 | SPECTRIN 95. |
| FT | REPEAT 2207 2243 | FT | REPEAT | 11981 | 12083 | SPECTRIN 96. |
| FT | REPEAT 2266 2302 | FT | REPEAT | 12083 | 12185 | SPECTRIN 97. |
| FT | REPEAT 2321 2357 | FT | REPEAT | 12185 | 12287 | SPECTRIN 98. |
| FT | REPEAT 2386 2422 | FT | REPEAT | 12287 | 12389 | SPECTRIN 99. |
| FT | REPEAT 2445 2481 | FT | REPEAT | 12389 | 12491 | SPECTRIN 100. |
| FT | REPEAT 2504 2540 | FT | REPEAT | 12491 | 12593 | SPECTRIN 101. |
| FT | REPEAT 2563 2599 | FT | REPEAT | 12593 | 12695 | SPECTRIN 102. |
| FT | REPEAT 2622 2658 | FT | REPEAT | 12695 | 12797 | SPECTRIN 103. |
| FT | REPEAT 2681 2717 | FT | REPEAT | 12797 | 12899 | SPECTRIN 104. |
| FT | REPEAT 2740 2776 | FT | REPEAT | 12899 | 12999 | SPECTRIN 105. |
| FT | REPEAT 2809 2845 | FT | REPEAT | 12999 | 13101 | SPECTRIN 106. |
| FT | REPEAT 2868 2904 | FT | REPEAT | 13101 | 13203 | SPECTRIN 107. |
| FT | REPEAT 2927 2963 | FT | REPEAT | 13203 | 13305 | SPECTRIN 108. |
| FT | REPEAT 2986 3022 | FT | REPEAT | 13305 | 13407 | SPECTRIN 109. |
| FT | REPEAT 3045 3081 | FT | REPEAT | 13407 | 13509 | SPECTRIN 110. |
| FT | REPEAT 3104 3140 | FT | REPEAT | 13509 | 13611 | SPECTRIN 111. |
| FT | REPEAT 3163 3199 | FT | REPEAT | 13611 | 13713 | SPECTRIN 112. |
| FT | REPEAT 3222 3258 | FT | REPEAT | 13713 | 13815 | SPECTRIN 113. |
| FT | REPEAT 3281 3317 | FT | REPEAT | 13815 | 13917 | SPECTRIN 114. |
| FT | REPEAT 3340 3376 | FT | REPEAT | 13917 | 14019 | SPECTRIN 115. |
| FT | REPEAT 3405 3441 | FT | REPEAT | 14019 | 14121 | SPECTRIN 116. |

Query Match 2.1%; Score 153; DB 1; Length 5938;

Best Local Similarity 17.6%; Pred. No. 4.4;

Matches 219; Conservative 175; Mismatches 412; Indels 438; Gaps 59;

264 RRADQVSDFTSGTSGYSSVPSACRHNSITQSSPEAPPSPRKRPPVDAIVPEAGHLI 323

4267 RHKDSMDEL-----FSHRSEIFGTGEGSQKTVL-----QKTESLIQYIAISLS 4311

324 ESPSGKRTTKSHI-DNQPLSSSSQGETSFSTVYESFP-----SSGGEGAIPESRSNGLAR 378

4312 NSERVARLERAQVLNQ-----FWETVEELSFWIEETRALIAQLPSAIDHEQLR 4361

379 SEE-----SARSQVQHAP-----VVAALRLNINWPKFKLWELAVAWETRLFMHCK 428

4362 QQQEEMRQLRESIAEKPHIDKLLKIGPOLKELNPEGEWEEK-----YQKAENMYAQIK 4417

429 VLEDESLGLKYPDSWSTARDVTDIWKTLRLDAPRGKPPFPKPPNDVFTVMTGNFESK 488

4418 EVRQKALAL--DEAVSQSTQITE-----FHDK-----IBPMLETLENL 4454

489 GSAVLSAVLDYNPD-----NSPTAPLYLVKPLMPEQCCR-----LIRFPQDR-- 534

4455 SSRLRMPPLIPAEVDKIRECISDNKSATVELEKLP--SFEALKRRGEELIGSQGADKDL 4513

535 -----PFE-----ILIPSPST----- 546

4514 AAKIQDKLDQWFFWEDIKARABEREIKFLDVLBAKFWYDMAALTITIKDTQDIVHD 4573

547 --SPSVPPVVSQPGAVEEVIQWLTMGQHS-----LVGRQWRAPFA-----KDAQYRKL 594

4574 LESPQIDPSIIKQVEAABTIKEETDGLHELEFIRILGAD--LIFACGETEKPEVRKSI 4631

595 REFQLEADPKPIIKERVHFFPAETGITRPDPFKRSVVPAEPEVQRTFK--VSQMLD 652

4632 DEMNNAWENLNTWKERLE-----KLEDAQAAVQYQDTIQAMFD 4671

653 WLLQLDNTWQPHLKLFPS-----RIQLGSKTYAIMTLRPHQIR-----HHKTDL 697

4672 W---LDNTV---IKLCTWPPVGTDLNTVKDQLNEMKBFKEVYQ--QIEMEKLNHQGL 4723

698 LSPSGTGEVNNQV-----GRMSRSVAKIRD-----VLGLGDVPSAVQGRFGSA 742

4724 MLKATDETDRDIIRPLELTTELKHELNWENLGEKTAHQHKLLEGALLAQGQHALEELMS-- 4781

743 KGMWVIDVDD-----TGDEWITETPSQAKWECDFVDKQHTLE--VRSVASEL--KS 791

4782 ---WLTHTELLDAQRPISGDPKVIIEVLAKHVLKNDVLAHQATVETVKNAGNELLESS 4838

792 AGNLQOLLVLEDRARDKVMQAIQIDRLINDLQROFSEQKHALNPVPEFRQWYVESYS 851

4839 AGDDASSL-----RSRLAEMNQWESVLQKTEEREQQLSTLQQAQHFHSEI--EDFL 4890

QY 852 RATRV5---HGRVPLAGLPD-----SQETLFLMN----- 880
 DB 4891 ELTRMESQASAKP-TGGLPETAREQLDTHMELYSQKAKEETYNQLLDKGRMLLSRDD 4949
 QY 881 --SGDPKQKQYQDIADWLQKRCDDTLKSLNIRVGRSAVIYMIADFWGLNEVHVHG 938
 DB 4950 SGSGSKTQSGVALLQKHVVSSKWEERKSL-----EEALN 4986
 QY 939 FSSKPRDEESP---TLSDCDVLVARSPAPPPSDIQ---RVRAVFKPELHSLKDVII-- 990
 DB 4987 LATERQNSLOQEPINWLTIAEQSLNIASPPSLILATVLSQIBEHKVFANEVNAHRDQIEL 5046
 QY 991 -----FSPKGVDPVLAKL-----SGDYDG-----DMAWVC 1016
 DB 5047 DOTGNQLKFLSKQDVVILKLLSVQSRWEKVVORSTIERGSLDDARKRAKQHEAW-- 5104
 QY 1017 WDPEIVDGFVNAEMPLEPDLRYLKDKKTTFKQLMASH-----GTGS 1058
 DB 5105 --KLIDWLEDAESHLDSLE--ISGDPKIKQLSKHKEFKTLGGKQPVYDTIRTGR 5160
 QY 1059 AAKEQTTY-DMTOKSFHALQBNFLGCMCTNYKERLCYINNVSNNKPAIILSSLVGNLVDQ 1117
 DB 5161 ALUKETLLPEDTQK-----LNFLEVRDKWDTVC-----GKSVR 5196
 QY 1118 S---KQGIYFNBAWAQLRRELGLGALSIPDPMYKSDSLWLGSGEPHIIIDYLFESIARPA 1174
 DB 5197 QHKLBEALLFS-----GQMDALQALVWLKVPQLAEDQPVH----- 5235
 QY 1175 IDKELEAFINAKAAKADTEDGAHFWDPDLASVYTFKISDKSRSSALLFTTLKRNIGEV 1234
 DB 5236 --GDLDLVNMLDAKH-----VFOKE-----LGRKRTGV 5262
 QY 1235 E--KEYGR-LVKNKEMRSDKDPYVVRVNOVYKWKCAITPEAMDK 1275
 DB 5263 QVLRKSGRELIENS--RDTTWKQQLQLSLRWDVCKLSVSK 5304
 RESULT 13
 TACC2 MOUSE STANDARD; PRT; 1035 AA.
 AC Q9JG6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transforming acidic coiled-coil-containing protein 2.
 GN TACC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
 RA Hashimoto K.;
 RT "Isolation of full-length cDNA clones from mouse brain cDNA library
 made by oligo-capping method."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May play a role in organizing centrosomal microtubules.
 CC May act as a tumor suppressor protein (By similarity).
 CC -1- SUBUNIT: Interacts with microtubules (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear; concentrated at centrosomes (By
 CC similarity).
 CC -1- SIMILARITY: Belongs to the TACC family.
 CC
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMEL; AB041546; BAA95031.1; --
 DR HSSP; P02649; 1LE4;
 DR MGJ; MG1:1928899; Tacc2.
 DR InterPro; IPR007707; TACC.
 DR Pfam; PF05010; TACC; 1.
 KW Coiled coil; Nuclear protein. POLY-PRO.
 FT DOMAIN 43 51
 FT DOMAIN 496 499 POLY-LYS.
 FT DOMAIN 763 791 COILED COIL (POTENTIAL).
 FT DOMAIN 834 1034 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1035 AA; 112715 MW; 0698015282F94C78 CRC64;
 Query Match 2.0%; Score 151; DB 1; Length 1035;
 Best Local Similarity 19.3%; Pred. No. 0.4;
 Matches 203; Conservative 127; Mismatches 368; Indels 354; Gaps 51;
 QY 71 SLNQARA-NFFTEAKAASNWVFKAHADPTLPWSKEPPRAATAGQOMALQTVLLEVLNR 129
 DB 6 SLQAPASDLNLEVSAMSSDSEAPETPESTTPVKAPPAPP----- 48
 QY 130 FMPPPNNITPQRTFGRTLSPGSLSRPTSTNTRKDEPANVTTFADPPKRLSTRSATGPPI- 188
 DB 49 --PPPEVTP-----EPEVIDPAPPEPGCISEPP-VVVPDGPSS--ESVEGSPR 94
 QY 189 --HGAIPKLPDPV-NTGSKRPSLES-----ENLNQCTKRAKGLSDNVAAPV 238
 DB 95 PSHSSAVFDEDEKFIASSGTYNLDFDSIELVDFNFSLEPCSDADSKGQC----- 143
 QY 239 PIASALDKVPTRRHANTRDPTATG-----HRRADQVDSFDTSQGTSGSSVFSACRHQ 292
 DB 144 -----KYSTRKSTESVPPSKSTLSLSLQASDFDGCSCGSPGAEAGTLTTDACGTGS 196
 QY 293 STTQSSFB-----APPS-----OPREKRPVDATVFEAGHLIESFKGRITKSHIDNQ 339
 DB 197 NSASSTLKRTKTRPPSLKKQATKPTETPPVKETQEPGE--ESPV---PSEELAPE 251
 QY 340 PLSSSSOGETSFSTYVESPPSSGGECAIPEPSRNSGLARSESARQOVHAPVVA--AR 397
 DB 252 TKTESATPEGAGCTLSDDTFLE--SPAVTATCPLTLEAEDV-----SLVSGGR 301
 QY 398 LRNIWPKFKWLHEAPLAVAVETRLFMCKVDLED-----BSLGLKYD-----PSWSTA 447
 DB 302 VQNSPPVGRK---SVPLTTASEAVEVTLSDSGQEDLPAGLSVRLFEFVSEDKGSWEQ 358
 QY 448 RDVT-----DIWKLRLDAFRGKPPPKP--PNDVFVTAMTGNFE 486
 DB 359 QENAPPTKIGKPKVAKMPLRRPKMTPEKLDNTPASP-PRSPTEPSTPIAKGTFTD 417
 QY 487 SKGSVVLSAVLDY--NPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFELIP--- 541
 DB 418 -----IDKWDDPNFNPFSSTKMQEGPKLSQ-----SYNFPDPACEESLDPFKA 462
 QY 542 -SPTSTSPVPVVKQPGAVEVIOQLTWGQSLVGRQWRAFFAKDAGYRKPLREFQLR 600
 DB 463 SSTPSPSPKSPASFTIPATTE-----ADGDGLNKPAK-----K 497
 QY 601 AEDPKPIKERVHFFAETGITFRPDVFKTRSVVPAEPEVQRTFVKVQMLDWLLQDNN 660
 DB 498 KKTPLKTMVDVMSVCSLFTFR--VKSKPSKPSLSDPPSQ----- 536
 QY 661 TWQPHLKLFSRIQLGSLKTYAIVMLPQHRIHHKHTLLSPSGTGEVMDGVGRMSRVAK 720
 DB 537 -----DP-----TPAATPE----- 545
 QY 721 RIRVDLGLGDPVAVQGRFGSAKGMWVIDVDGDEDIETYSQRKWECDVD----- 774
 DB 546 -----APSAIS-----TVVHATDEEKAV--TSQKWTCTMTDLDADKQ 581
 QY 775 KQHTLEVRVASELKSAGLNQLLPVLEDRADKVKVROAIGDRLINDLQRFSEKHA 834
 DB 582 DFPQPSDLNFVNETKFNPSSEE-----LDYRNSYEIYMEKLGSSLPQD---DTPAKQA 634

```
QY 835 L-----NRPE-----FRQWVYESYSSRATRVSHGRVFPFLAGLDP 869
DQ 635 LVLMDFTQSPVSPKPPVRMSDSPTPCGSSSPED--TEALVNAATKLQH---PVARGLPS 689
QY 870 SOEETINFLMNSGFPKPK--QKYLQDIANDLQKR-----KCDTLKSLNIRVG 915
DQ 690 SOEPLQV-----PEKPSQKLEAMALGTAPAEIETAPGAFASATLTLRL----- 737
QY 916 RGAITYMIADFQVLENEVHVH--FSGKFRDEESETLLSDCDVLVARSPA---HPPS 969
DQ 738 --AHPASLCGALGYLEPDLAEKNPPVFAQKLOEELLE-FAVWHIEALKLARQIALASRSRQ 794
QY 970 DQIRVRAVEKPELHSLKDVIIESTKGDVPLAK 1001
DQ 795 DTKR-EAHPDPV-SISKALYSRICSTEVK 824

RESULT 14
MINT_MOUSE
ID MINT_MOUSE STANDARD; PRT; 3644 AA.
AC Q62504; Q8OTN9; Q9P84; Q9QW2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
TISSUE=Testis;
RX MEDLINE=99379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel mxx2 binding protein, recognizes and
RT regulates the rat osteocalcin promoter.";
RL Biochemistry 38:10678-10690(1999).
RN [2]
SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
RA Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA Watanhiki A., Watanabe M., Hayashizaki Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawauchi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT J/Su(H) activity.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 318-578 FROM N.A.
TISSUE=Cochlea;
RC MEDLINE=97237053; PubMed=9119401;
RA Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fizes C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341 (1997).
RN [5]
SEQUENCE OF 2598-3644 FROM N.A.
TISSUE=Brain;
RC MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
```

```
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
RN [6]
TISSUE SPECIFICITY.
RP MEDLINE=2261914; PubMed=12374742;
RX Oswald F., Kostecka U., Astrahantseff K., Bourteelle S., Dallingier K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway.";
RL EMBL J. 21:5417-5426(2002).
RN [7]
FUNCTION, AND TISSUE SPECIFICITY.
RP MEDLINE=22483652; PubMed=12594956;
RX Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT "Regulation of marginal zone B cell development by MINT, a suppressor
RL of Notch/RBP-J signaling pathway.";
CC Immunity 18:301-312(2003).
CC -I- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative
CC regulator of the Notch pathway via its interaction with RBPJH,
CC which prevents the association between NOTCH1 and RBPJH, and
CC therefore suppresses the transactivation activity of Notch
CC signaling. Blocks the differentiation of precursor B cells into
CC marginal zone B cells. Probably represses transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins. May bind both to DNA and RNA.
CC -I- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and
CC MTAL1. Interacts with the nuclear receptors RAR and RXR.
CC Interacts with RAR in absence of ligand. Bind to the steroid
CC receptor RNA coactivator SRA (By similarity). Interacts with MSX2.
CC Interacts with RBPJH; this interaction may prevent the
CC interaction between RBPJH and NOTCH1.
CC -I- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62504-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q62504-2; Sequence=VSP_008564;
CC Note=No experimental confirmation available;
CC -I- TISSUE SPECIFICITY: Highly expressed in testis. Expressed at lower
CC level in brain, lung, spleen, liver and kidney. Weakly expressed
CC in cardiac and skeletal muscles and ovary. In spleen, it is
CC expressed in follicular B-cells, while it is weakly expressed in
CC marginal zone B-cells.
CC -I- DOMAIN: The RID domain mediates the interaction with nuclear
CC receptors.
CC -I- DOMAIN: The SPOC domain, which mediates the interaction with
CC NCOR2, is essential for the repressive activity (By similarity).
CC -I- SIMILARITY: Belongs to the Spen family.
CC -I- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC -I- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -I- SIMILARITY: Contains 1 SPOC domain.
CC -I- CAUTION: Ref.4 sequence differs from that shown due to multiple
CC frameshifts and conflicts that create stop codons.
CC -I- CAUTION: Ref.5 sequence differs from that shown due to what seems
CC to be the presence of intronic sequence in the cDNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; B726481; -; NOT ANNOTATED_CDS.
CC EMBL; AF156529; AAD55931.1; ALT_INIT.
DR
```

DR EMBL; AB055980; BAB32786.1; --
 DR EMBL; 278150; CAB01562.1; ALT_SEQ.
 DR EMBL; AX12402; BAC65684.2; ALT_SEQ.
 DR MGD; MGI:1891706; Mint.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR SMART; SM00360; RRM; 3.
 DR SMART; PF00076; rrm; 3.
 DR PROSITE; PS0102; RRM; 4.
 DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 DR PROSITE; PS0917; SPOC; 1.
 KW Transcription regulation; Repressor; Nuclear protein; DNA-binding;
 KW RNA-binding; Repeat; Coiled coil; Alternative splicing; Polymorphism.
 FT DOMAIN 1 574
 FT 1 574
 FT 6 81
 FT 336 416
 FT 439 514
 FT 518 590
 FT 559 575
 FT 822 850
 FT 1185 1206
 FT 1509 1544
 FT 1607 1627
 FT 2216 2704
 FT 3478 3644
 FT 2138 2462
 FT 2706 2845
 FT 125 277
 FT 236 326
 FT 648 721
 FT 702 832
 FT 2101 2233
 FT 2377 2518
 FT 2950 3475
 FT 618 640
 FT VARSPLIC
 FT 348 348
 FT 762 762
 FT 773 773
 FT 933 933
 FT 754 754
 FT 1524 1524
 FT 1560 1560
 FT 1570 1570
 FT 1574 1574
 FT 1609 1609
 FT 1659 1659
 FT 1669 1669
 FT 1705 1705
 FT 1815 1815
 FT 2097 2097
 FT 2201 2202
 FT 2322 2322
 FT 2385 2385
 FT 2502 2502
 FT 2505 2505
 FT 2519 2519
 FT 2554 2554
 FT 2678 2688
 FT 3010 3010
 FT 3644 3644
 SEQUENCE
 Query Match 2.0%; Score 151; DB 1; Length 3644;
 Best Local Similarity 21.0%; Pred. No. 2.8;
 Matches 137; Conservative 63; Mismatches 254; Indels 198; Gaps 29;
 QY 3 PITPRKNSVETIINRLNNDYNLGLQCVAADTLTPHRRKELAESDEDFGRHDKIYRALN 62
 DB 1963 PTPPRGRPP--KTRRAED-----GEHERKEPAETPRPREG----- 1998
 QY 63 FLVWRKDDSLNQ-----EANFFIEAKASGNWVPKAAHDDPTLPWSKE 106
 DB 1999 ---WRSPRSQKAAAGPQGRGRNEQKVBAAGAAQAQASTREGNPKSRGEREAASEPKR 2055

107 PRAATAGQWALQTVLLEVLNRFMPNNPTGRTTSGPGLSRPTSTNTRKDEP 166
 2056 DRDPSIDKS-GPDTPFEVLER--XPEKTYKSKGRARSTESGVDRAHQSLWAAR 2112
 QY 167 ANVTADPPKRSLTRSATGPIHGAIAIPKFPDVPNTGSKRPSL-----ESEN 215
 DB 2113 AAGQAAD-----KEAG-----AAASPQSESPQKSGSSPOLANPADPDEAEESA 2161
 QY 216 NQCTKRAKG-----KLSNVAATAAPVPIASALDKVPTRRHANTROPTAT----- 261
 DB 2162 SASATPEGTQARQIELEQAVQNIAPKLPSSAAAASK-----GTATATAASEPA 2214
 QY 262 ---GHRADVDSPDTSGTSYSSVFSACRNQSTQSSFEAPPS-----QPRE--- 308
 DB 2215 PERGHKPAHQAS--ETELAAAGSIISDA-----SGEPENFSAPPSPFGSQTHREGME 2267
 QY 309 --KRPVDATFEAGHLIESPKGRRTTKSHIDNQLSSSQGSETSFSTYVESPPSSGEGA 366
 DB 2268 PGLHEAESGILETGATATESA---PQVSALD--PPEGSADTKETRGNSGVQEAQSKAE 2323
 QY 367 IPEPSRNGHARSEARSQVQVHAPVVAARLENINWPKWLHEAPLAVANEVTLFMH 426
 DB 2324 VTPPRKDKG--RQKTRRRKENANKVVA-----ITETRAS 2357
 QY 427 CKVDLEDESGLKXKVDPSWSTARDVTDINKLYLDAFRGKPPPEKBPND---VFVTAMT 482
 DB 2358 EASQTQSES-----PAEAEATAATP-----EAPQEKSPKPPSPACCTDFPSKT 2403
 QY 483 GNFEKGSAYVLGAVLDYNDPNSFTAPLVKLKPLMFEGGCLTRRFGDRPFBI--LI 540
 DB 2404 PPAESLSQE--NSAAEKTCKAPVLPAFLPALMDD-----GPQAREKVVHSII 2452
 QY 541 PSPISISPS-----VPPVYSKQGVAV-----EEVIGWLT 569
 DB 2453 ESDPVTFPSPDSGIPPTPIPLVTIAKUPPVVI---PGGVHQSPPPKTEWIT 2501

RESULT 15
 SMC4 YEAST
 ID SMC4 YEAST STANDARD; PRT; 1418 AA.
 AC Q12267;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Structural maintenance of chromosome 4.
 GN SMC4 OR YLR086W OR L9449.5.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RC MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
 RA Benes V., Brueckner M., Deltus H., Dubois E., Duesterhoeft A.,
 RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
 RA Louis E.J., Messenguy F., Mewes H.-W., Miesga T., Moestl D.,
 RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
 RA Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,
 RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
 RA Underwood A.P., Urrestazu A., Vandenbol M., Verhasselt P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohnseil J.D.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
 RL Nature 387:87-90(1997).
 RN [2]
 RP IDENTIFICATION IN A CONDENSIN COMPLEX WITH SMC2; BRN1; YCS4 AND
 RP YCG1.
 RX MEDLINE=20273907; PubMed=10811823;
 RA Freeman L., Aragon-Alcaide L., Strumnikov A.;
 RT "The condensin complex governs chromosome condensation and mitotic

| Result No. | Query % | | DB | ID | Description |
|------------|---------|-------|------|----|-------------|
| | Score | Match | | | |
| 1 | 7397 | 100.0 | 1402 | 3 | Q9Y7G6 |
| 2 | 328 | 4.4 | 1116 | 10 | Q9ZR77 |
| 3 | 328 | 4.4 | 1136 | 10 | Q9SG32 |
| 4 | 326 | 4.4 | 1196 | 10 | Q9LKP0 |
| 5 | 325 | 4.4 | 1218 | 10 | Q8LKH9 |
| 6 | 322.5 | 4.4 | 775 | 10 | Q9ZRK3 |
| 7 | 320.5 | 4.3 | 1136 | 10 | Q7XMX1 |
| 8 | 312 | 4.2 | 1351 | 3 | Q87253 |
| 9 | 298 | 4.0 | 1964 | 5 | Q95ZG6 |
| 10 | 297 | 4.0 | 1114 | 10 | Q9ZR58 |
| 11 | 296.5 | 4.0 | 1780 | 5 | Q19285 |
| 12 | 293.5 | 4.0 | 1107 | 10 | Q9LQW2 |
| 13 | 294.5 | 4.0 | 1484 | 3 | Q8TGV4 |
| 14 | 293.5 | 4.0 | 1883 | 5 | Q95ZG7 |
| 15 | 287 | 3.9 | 1107 | 10 | Q8H1K9 |
| 16 | 285.5 | 3.9 | 513 | 10 | Q9ZRK5 |

```
QY 121 TVLLEVLNRFMPBPNNTPGRTTSGPSGLSRPTSTNTKEKDEPANVTADPEKSLT 180
Db 121 TVLLEVLNRFMPBPNNTPGRTTSGPSGLSRPTSTNTKEKDEPANVTADPEKSLT 180
QY 181 RSATGPIHGAAPLKPDPVNTGSKRPSLESENINQCTKRAKGLSDNVAAPVPI 240
Db 181 RSATGPIHGAAPLKPDPVNTGSKRPSLESENINQCTKRAKGLSDNVAAPVPI 240
QY 241 ASALDKVPTERRANTDPATGHRADQVSDFTSGTSGSYSSVSACRHNQSTTQSSFE 300
Db 241 ASALDKVPTERRANTDPATGHRADQVSDFTSGTSGSYSSVSACRHNQSTTQSSFE 300
QY 301 APPSQPREKRPVDATVEAGHLIESPKGRTTKSHIDNQPLSSSSQGETSFSTYYESPPS 360
Db 301 APPSQPREKRPVDATVEAGHLIESPKGRTTKSHIDNQPLSSSSQGETSFSTYYESPPS 360
QY 361 SGGECAIPSPSRNGLARSEESARSQVHVAVVAARLNTWPKPKWLHEAPLAVAVEV 420
Db 361 SGGECAIPSPSRNGLARSEESARSQVHVAVVAARLNTWPKPKWLHEAPLAVAVEV 420
QY 421 TRLFMECKVDLEDESILGLKYDPSWSTARDVTDIWTKLYALDAFRGKPFPEKPPNDVFVTA 480
Db 421 TRLFMECKVDLEDESILGLKYDPSWSTARDVTDIWTKLYALDAFRGKPFPEKPPNDVFVTA 480
QY 481 MTGNFESKGSAVVLSAVLDYNDPNSPTAPLYLVKLPMPFEGCRLTRFGPDREFFILI 540
Db 481 MTGNFESKGSAVVLSAVLDYNDPNSPTAPLYLVKLPMPFEGCRLTRFGPDREFFILI 540
QY 541 PPSPTSPSPVVPVSKQGAVEVIOWLTMGHSILVGROWRAFFAKDAGYKRPREFOLR 600
Db 541 PPSPTSPSPVVPVSKQGAVEVIOWLTMGHSILVGROWRAFFAKDAGYKRPREFOLR 600
QY 601 AEDKPIIKERHFFAETGITFRPDVFKTRSVVPAEPEVEQTEFKVQSMQLDILLQDNN 660
Db 601 AEDKPIIKERHFFAETGITFRPDVFKTRSVVPAEPEVEQTEFKVQSMQLDILLQDNN 660
QY 661 TWQPHLKLSRIQLGSKTYAIVMLEPHQIRHKTDLTLLSPSGTGEVMDGVGRMSRVAK 720
Db 661 TWQPHLKLSRIQLGSKTYAIVMLEPHQIRHKTDLTLLSPSGTGEVMDGVGRMSRVAK 720
QY 721 RIRDVLGLGVPVSAVQGRFGSAGMMVIVDDTGDDEDTETVPSQRKKECDFVDKHQRTL 780
Db 721 RIRDVLGLGVPVSAVQGRFGSAGMMVIVDDTGDDEDTETVPSQRKKECDFVDKHQRTL 780
QY 781 EVRSVASELSKAGNLQLLPVEDRARDKVRQAIGDLINDLQORSEQKHALNRVPE 840
Db 781 EVRSVASELSKAGNLQLLPVEDRARDKVRQAIGDLINDLQORSEQKHALNRVPE 840
QY 841 FRQWVYESYSSRATRVSHGRVPPFVPLAGLPDSQEBTLNFMNSGFDPKKQYLQDIAWDLQK 900
Db 841 FRQWVYESYSSRATRVSHGRVPPFVPLAGLPDSQEBTLNFMNSGFDPKKQYLQDIAWDLQK 900
QY 901 RKCDTLKSLNIRVGSAYIYMTADFWGLVLENEHVHVGFSKFRDEESFTLLSDCVILV 960
Db 901 RKCDTLKSLNIRVGSAYIYMTADFWGLVLENEHVHVGFSKFRDEESFTLLSDCVILV 960
QY 961 ARSPAHFPDSDIQVRVAFKPELHSLKDVIIFSTKGDVPLAKLSGGDYDGDMAWVWCPDE 1020
Db 961 ARSPAHFPDSDIQVRVAFKPELHSLKDVIIFSTKGDVPLAKLSGGDYDGDMAWVWCPDE 1020
QY 1021 IVDFGVNAEMPLSPDISRYIKDKTTFKQLMASHGTGSAKQCTTVDMTQKSFHALQDN 1080
Db 1021 IVDFGVNAEMPLSPDISRYIKDKTTFKQLMASHGTGSAKQCTTVDMTQKSFHALQDN 1080
QY 1081 FLGWCNTYKRELVCYNNVSNKPAIILSLVGNLVQSKQGIIVFNEASWAQLRRELLGGA 1140
Db 1081 FLGWCNTYKRELVCYNNVSNKPAIILSLVGNLVQSKQGIIVFNEASWAQLRRELLGGA 1140
QY 1141 LSLPDPMYKSDSNLGRGEPTTHIDYLFKSTARPAIDKELEAFHNAMKAADTGDGAHFW 1200
Db 1141 LSLPDPMYKSDSNLGRGEPTTHIDYLFKSTARPAIDKELEAFHNAMKAADTGDGAHFW 1200
QY 1201 PDLASYTYTFPEKISDKSRSSALLFTTLKNRIGEBVEKEYGRLVKNKEMRSDKDPYPRVNVQ 1260
```

```
Db 1201 PDLASYTYTFPEKISDKSRSSALLFTTLKNRIGEBVEKEYGRLVKNKEMRSDKDPYPRVNVQ 1260
QY 1261 VYKWCATPEAMDKSGANYDSKVIILLLESLFLADREMTWALLRASTAFKLYYHKSXP 1320
Db 1261 VYKWCATPEAMDKSGANYDSKVIILLLESLFLADREMTWALLRASTAFKLYYHKSXP 1320
QY 1321 VWQWAGRLQALYIKAOQTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDSYPPDP 1380
Db 1321 VWQWAGRLQALYIKAOQTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDSYPPDP 1380
QY 1381 EYEVGLGDDDDFDGIGTGNGDY 1402
Db 1381 EYEVGLGDDDDFDGIGTGNGDY 1402
RESULT 2
Q9ZRY7 PRELIMINARY; PRT; 1116 AA.
AC Q9ZRY7;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE RNA-directed RNA polymerase.
GN RUP.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamiales; Solanales; Solanaceae; Nicotiana.
CX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI; TISSUE=Leaf;
RA Schiebel W., Pelissier T., Riedel L., Thalmair S., Schiebel R.,
RA Kempe D., Lottspeich F., Saenger H.L., Wassenecker M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT Tomato.";
RI Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011576; CAA09697.1;
DR PIR; T30828; T30828.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR InterPro; IPR007855; RdRP.
DR Pfam; PF05183; RdRP; 1.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1116 AA; 127553 MW; 8D9B89DA826D31DB CRC64;
```

```
Query Match 4.4%; Score 328; DB 10; Length 1116;
Best Local Similarity 19.3%; Pred. No. 2.9e-13;
Matches 211; Conservative 149; Mismatches 386; Indels 314; Gaps 42;
QY 441 DPSNSTARDVTDIWKTL-----YRLDAFRGKPFPEKPPNDVFFV--TAMTGNFES 487
Db 193 DDQWVRTTDTFSPWIGLSSSLCLEFRNGVQLPNFSSFFYKESMNQFILQTGTFSSQ 252
QY 488 KGSVULSAVLNDYNDPNSPTAPLYLVKLPMPFEGC--RLTRFGPDREFFILIPTS 545
Db 253 KLALVPIV----HLPEGIELPYKILFKISLLI--QHGFPGGLANF---NFFQLVDPRRN 304
QY 546 TSPSPVPPVSKQGAVEV-----IQWLT-MQCHSLVGRQ----- 579
Db 305 -----IACIEHALEKLYLKECCYDPRMLTEQYDEYLGKRLPKSPITLDDGLVY 356
QY 580 -----WRAFFAKDAGYKRPREFOLR--AEDPKFIK-----ERVHFF-----A 616
Db 357 VRRVVVTFCKYVF---CGPEVNVSNRVLNRVSEIDNNFLRVFVDEWEKIHSTOLLFRA 413
QY 617 ETGITPDPVFK-----TRSVPAEPEVEQTEFKVQSMQLD---W-----LLQDNN 661
Db 414 STNGTDTDIYERILSTRNGFIIGDKRFEFLAFSSQLADNSVWMPASRPLTANDIRT 473
QY 662 WQPHLKLSRI-----QLGSKTYAIVMLEPHQIRHKTDL-----SPSGTGEVMDGV 711
Db 474 WMGDFRQIRNVAKYVARLGOSFSSRETL---SVGRHEVEVFDVACSLHGTNYIFSDGI 530
```

QY 712 GMSRSVAKRIKRDVGLGDPVSAVQVGRFGSAKGMWVIDVDGTDGDEDTIETVPSQKWECD 771
 Db 531 GKISADFAHRTIKGLQYTSFIRYGGYGVAVD-----PYSSMK----- 574
 QY 772 FVDKQRTLEVRASVASELKSAGLQLL-----PVLEDRADKVKMRQAIGDRLINDLQ 825
 Db 575 -----LSLRKSLMKYVESNNIKLDVLGWSKYQCYLNRQLVTLSTLGLVIDVLEQKQ 626
 QY 826 RQFSQKHALNRPVEFRQWVYVESYSSRATRVSHGRVPELAGLPDSQEBTLNFMNSGDP 885
 Db 627 NEAVQDLDAIHLSKAEALMS-----PGENTILKEMLNCYMP 669
 QY 886 KXQKQLOPIADLQKCKDTLKSKLNRVGRSAVIYMIADFWGLVENEHVGFSS----- 941
 Db 670 DAEPFLSNMLQTFRASKLLDKTRTRIFIPNGRTWMCGLDESRTLEYQGVQVPSGAGR 729
 QY 942 KPRDEEFP--TLSDSCD-----VLVARSAPHPSDIORVAVFKPELHSLKDVILFEST 993
 Db 730 QFESHFPNDSGANSDFILKGNVVVAKNPLCPHGDIVLRVADVPAHHHMDVCPVFPQ 789
 QY 994 KGDVPLAKKLGGDYDGMWVCMWDPDPIVDGFWNAEMPLEDPLSRYLKOKTTFKQLMAS 1053
 Db 790 KGRPHPNESGSLDGDIFVFCWDPDLIPPROVQSMYTPATTTQLDHD----- 839
 QY 1054 HGTGSAKXEQITYDMIOKSPHALQPNFLGCTNYKERLCVINNSV----- 1100
 Db 840 -----VTIEVEEYF-----TNY-----TINDSLGIIANAHVVFADRE 872
 QY 1101 -----NKPAIILSSVLGNLDOSKOGIVFNEASWAQLRELLGGALSPLD--PMYKSPSWL 1154
 Db 873 PDAMSDDCKQLAQLFSTAVDPFKGVPAEIPS--QLRPKPYDPFMKPKDPTYPEKVI 930
 QY 1155 GRGETHIIDYLFKSIAPDAIDKELEAFHNAWKAATKOTEDGAHWPDPDLASYTFFKES 1214
 Db 931 GK-----LFQVKVKN-----TPQASSIATFTEDVA 955
 QY 1215 DKSRSSALLFTTLKNRIGEV--EKEYGRIVKNKEMRDSKDPYVRYNQVYKWKCAITPE 1271
 Db 956 KKSYSOSDMVDFEDYIDFAFYKSEY-----DNKLGIMDYIGKITE-----AEILSG 1004
 QY 1272 AMDKSGANYDSK-----VIRILELSFLADRENTW-----ALPRASTAFKLYY 1314
 Db 1005 GIMKASKTFDRKDAEAGVAVRCLR-----XEARAWFKRRSDIDDMMLAKASAWHYVTY 1058
 QY 1315 HXS-----PKFVQWAGRLAVIKAQWTSRP 1340
 Db 1059 HTYNGLYNEGLKRDHF--SFPKCVVDQLIQIKKAKARKEP 1098
 RESULT 3
 Q9SG02 PRELIMINARY; PRT; 1196 AA.
 AC Q9SG02; PRELIMINARY; PRT; 1196 AA.
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Putative RNA-directed RNA polymerase.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ev. Columbia;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC t1g12 genomic sequence."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
 RA Mewes H.W., Lencke X., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ev. Columbia;
 RX MEDLINE=20306668; PubMed=10850495;
 RA Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,
 RA Morel J.-B., Jouette D., Lacombe A.-M., Nikic S., Picault N.,
 RA Ramoue K., Sanjal M., Vo T.-A., Vaucheret H.;
 RT "Arabidopsis SGS2 and SGS3 Genes are Required for Posttranscriptional
 RT Gene Silencing and Natural Virus Resistance."
 RL Cell 101:533-542(2000).
 DR EMBL; AC013229; AAG52184.1; --
 DR EMBL; AL132964; CAB71285.1; --
 DR EMBL; AP239718; AAF73959.1; --
 DR GO; GO:000368; F.RNA-directed RNA polymerase activity; IEA.
 DR InterPro; IPR007855; RDRP.
 DR Pfam; PF05183; RDRP; 1.
 KW RNA-directed RNA polymerase.
 SQ SEQUENCE 1196 AA; 136927 MW; 812DBB9CDC82C9F CRC64;
 Query Match 4.4%; Score 328; DB 10; Length 1196;
 Best Local Similarity 22.7%; Pred. No. 3.2e-13;
 Matches 197; Conservative 126; Mismatches 312; Indels 234; Gaps 42;
 QY 535 FFEILPSTSTSPSPVPPVSKQGVAVBEVIOWLTMGQHSLVGRQWRAFAKDAKVKPL 594
 Db 466 FVAPIVKDLTSSFSQKTYVFKRKS-----LTDG-FKLCGRKY-SFLAFSAN----- 512
 QY 595 REFQRAEDPKPIKERVHFFAETGTFPRDFKRSVVPVPAEPEQRTPEFVSOMLDWL 654
 Db 513 ---QLR-----DRSAWFFAEDG-----KTR-----VSDIKTM 537
 QY 655 LQI-DNNTWQPHLKFRIQLGLSKTYAINTLEPHQIRHKTDLTSPSTGEVMDGVGR 713
 Db 538 GKFDKNV---AKCAARMGLCFSTYATVDMVPEV---DTEVPDIERNGVFSDIGT 590
 QY 714 MSRSVAKRTRDVLGLGDV---PSAVQGRGSAKGMWVIDVDGTDGDEDTIETVPSQ----- 765
 Db 591 ITPLDADEVMEKLL-DVHYSPCAVOIRYAGFGV-----VARPFSKDDGTR 636
 QY 766 ---RKWECDFVQKHQSTLEVRVASELSAGLNLQLPVLEDRADKVKMRQAIGDRLIN 822
 Db 637 LALRDSKKFFSKH-TILEICS-WIRFQPGFLNRQIITLS-----VLGVDPDEIFW 685
 QY 823 DLQRFSEOKHALNRPVEFRQWVYVESYSSRATRVSHGRVPELAGLPDSQEBTLNFMNSG 882
 Db 686 DMQ---ESMLYKLNRIILDTDVAFEVLTAACA-----EGNTAAINLSAG 727
 QY 883 FDPKKQYLQDIADWLQKRCQDTLKSKLNRVGRSAVIYMIADFWGLVENEHVGFSS-- 940
 Db 728 FKPTSPHLRGMSSVRIAQLWGLREKSLFVTSGRWLMGLDEAGILEHGCQFIQVSKP 787
 QY 941 -----SKPRDEESFTLLSDCVLVARSAPHPSDIORVAVFKPELHSLKDVIL 990
 Db 788 SIENCFSKHGRSPKTKTDLVVKGY-VAIAKNPCLHPGDVRLILEADVLPQLHMYDCHI 846
 QY 991 FSTKGDVPLAKKLGGDYDGMWVCMWDPDPIVDGFWNAEMPLEDPLSRY--LKQKTKTF 1048
 Db 847 FPKGDRPHNTNAGSGLDGLYFVANDQKLI-----PPNKSYPAMHYDAEEK 896
 QY 1049 QLMASHGTGSAAKEQTYTDMIOKSFHALQPNFLGCTN-----YKERLCVINNSVSKFA 1104
 Db 897 SL-----GRAVNHQDIIDFFARN-----LANEGLGTICNAHVHADRSY---GAMDEBC 943
 QY 1105 IILSSVLGNLDOSKOGIVFNEASWAQLRELLGGALSPLD--PMYKSDSWLGR----- 1156
 Db 944 LLALAEATAVDPFKTKGVISMP--PHLKPPLYDPFMGKHEDQTYKSKILGRILRVKKE 1001
 QY 1157 -----GEPTHI-----IDYLFKSIAP-----AIDKELEAFHNAWKAAD 1191


```

Db 1002 VYDEDAEASSESTDPSAIPYDAVLEIPGFEDLIPEAWGHKCLYDQGLIGLLGOYKVOKE 1061
QY 1192 TE-DGAEWDPDLASYYTFEKEISDKSRSSALLFTTLKNRIGEVKEYGRVLVKNKEMRDS 1250
Db 1062 BEIVTGHWS---MPKYTSKKQGLKER-----LXHSYNSLKKEF-RKVFEETIPDH 1109
QY 1251 KDPYPRVNVQVEK-----WCAIT--PEAMDKSGANYDSKVRLLELSFLADRENMTWALL 1304
Db 1110 ENLSBEEXNILEKASAWHYVTHPEWVKKS-----LELQ---DPDESSHAAM 1155
QY 1305 RASTAFKLYYHKSPPFVQWAGROLAYIK 1333
Db 1156 LS-----FAW-IAADYLARIK 1170

RESULT 4
Q9LKP0 PRELIMINARY; PRT; 1196 AA.
AC Q9LKP0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RNA-dependent RNA polymerase.
GN SDE1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20306669; PubMed=10850496;
RA Dalmay T., Hamilton A., Rudd S., Angell S., Baulcombe D.C.;
RT "An RNA-dependent RNA polymerase gene in Arabidopsis is required for
RT posttranscriptional gene silencing mediated by a transgene but not by
RT a virus."
RL Cell 101:543-553 (2000).
DR EMBL; AF268093; AAF74208.1; -
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR InterPro; IPR007855; RDRP.
DR Pfam; PF05183; RDRP; 1.
DR RNA-directed RNA polymerase.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136747 MW; C1B1961CDC11F708 CRC64;

```

```

Query Match 4.4%; Score 326; DB 10; Length 1196;
Best Local Similarity 22.5%; Pred. No. 4.4e-13;
Matches 200; Conservative 119; Mismatches 297; Indels 272; Gaps 42;

QY 535 PFELIPSTSTSPSPVVPVSKPGAVEVQIQLTWGQSLVGRQWRAFFAKDAGRKPL 594
Db 466 FVAPIVKDLTSSFSQKTYVFRKVKSI-----LTDG-FKLCGRKY-SFLASAN----- 512
QY 595 REFQRAEDPKPIKERVHFFARTGITFRPDVFKTRSVVPABEPVEQRTPEFKVQSLDWL 654
Db 513 ---QLR-----DRSAMPFADG-----KTR-----VSDIKTWM 537
QY 655 LQL-DNNTWOPHLKPSIOLGLSKTYAMTLEPHQIRHHKHTDLLSPSGTGVMDGVGR 713
Db 538 GKFKDKNV-----AKCAARMGLCFSSYATVDVMPHEV-----DTEVPDIERNGVFSDIGIT 590
QY 714 MSRSVAKRIIDVLGLGV---PSAVQGRFGSAGXGMWVIDVDDTGDGDWETIYPSQKWEK 770
Db 591 ITPDLAGVNEKUL-DVHVSFCAYQIRVAGFKGV-----VARWPS----- 630
QY 771 DFVDKHQRTLEVRSAVELSKAGLNLQLPVLIEDRDKVKVQKQAOAGDRLND----- 823
Db 631 -----KSDGIRLAL-----RD--SMKKIPSKHTILEICSWTRF 661
QY 824 ----LQRFSEQXHALNRPVEFFQWVYESYSSRATRV---SHGRVFFLAGLPDSQDETUN 876
Db 662 QPGFLNKQITLISLVGPVIEIFWDMQESMLYKLDRLIVDTVAFEVLTAACAEQNTAA 721
QY 877 FLAMSGFDPKQKYLQDIANDLQKRKCDTLKSLNTRVGRSAIYIMIADFWGLVENEVH 936

```

```

Db 722 IMLSAGFKPTEPHLRGMLSSVRIAQMLGRKSKSIFVTSGRWLMGCCDEAGILHGGCF 781
QY 937 VGFS-----SKFRDEEESFTLLSDCDVIVAKSPAHFPSDIORVRAVFPFELHS 984
Db 782 IQVSKPSIENCFSKGRSFKETKQLEVVVKGY-VAIAKNPCLPHPGVRILEAVDPQLHH 840
QY 985 LKDVIIIFSTKGVPLAKLKSGLDGMVWQWDEIIVDGVFVNAEMPLEPDLRSY--LKK 1042
Db 841 MYDCLIFPGQGRPHPTNEASGSDLDGLYFVAVDQKLI-----PPNKSYPAMHY 890
QY 1043 DKITTFQOLMASHGTGSAKEQTTYDMIQKSFHFAQPNFLGMCNTN----YKERLCYINNS 1098
Db 891 DRAEEKSL-----GRAVNHQIIDIFFAEN-----LANEQLTICNAHVVHADRSY---G 937
QY 1099 VSNKPAIILSSLVGNLDQSKOGIVFNEASQAQLRRELGGALSPLD-PMYKSDSWLGRG 1157
Db 938 AMDEECILLAEALAAATAVDFPKTKIVSMP--PHLKPLYPDFMGKEDYQTYKSKNLGR- 994
QY 1158 EPTHIIIDYLFKSIARP---AIDKELEA-----AHFWDPDLASYTFEKEISDKSRSSALLFTTLKNRI 1231
Db 995 -----LYRFPKVEYDEDAEASSEESTDPSAIPYDAVLEIPGFEDLIPEAWGHK 1042
QY 1191 DTEDG-----AHFWDPDLASYTFEKEISDKSRSSALLFTTLKNRI 1231
Db 1043 CSYDGLQILGLGOYKVOKEEIVTGHWS---MPKYTSKKQGLKER-----LXHSY 1091
QY 1232 GEVEKEVGRVLVKNKEMRDSKDPYPRVNVQVEK---WCAIT--PEAMDKSGANYDSKVI 1285
Db 1092 NSLKKEF-RKVFEETIPDHENLSEBEKNILEKASAWHYVTHPEWVKKS----- 1141
QY 1286 RLLELSFLADRENMTWALLRASTAFKLYYHKSPPFVQWAGROLAYIK 1333
Db 1142 --LELQ---DPDESSHAAMLS-----FAW-IAADYLARIK 1170

RESULT 5
Q9LHH9 PRELIMINARY; PRT; 1218 AA.
ID Q9LHH9
AC Q9LHH9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative RNA-directed RNA polymerase.
GN B1074C08.21.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
RT clone:B1074C08."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004357; BAC00725.1; -
DR Gramene; Q8LHH9; -
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR InterPro; IPR007855; RDRP.
DR Pfam; PF05183; RDRP; 1.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1218 AA; 136533 MW; 984FD38DC5474482 CRC64;

Query Match 4.4%; Score 325; DB 10; Length 1218;
Best Local Similarity 21.6%; Pred. No. 5.3e-13;
Matches 199; Conservative 149; Mismatches 318; Indels 254; Gaps 43;

QY 497 VLDYNPNDSPTAPLYLVKLKPLMPFEOQCRLTRRPGDRPFELLIPSTSTSPSPVPSK 556
Db 482 VLNNVNLNSPTAPI-----VKDLM-----SNFFQ-----Q 506

```

| | | | |
|----|------|--|------|
| QY | 557 | QPCAVEVQIQLTMQHSIVGQWRAFFAKDAGYKRLREFQLRADPKPIKERVHFFA | 616 |
| Db | 507 | KTTAVKRVEMLLTEGPH-MCGRKY-SFLAFSSN-----QLR | 548 |
| QY | 617 | ETGIFRPDPVFKTRSVVPAEPEVEQRTKFKVQSMQMLDNLQLDNNNTWQHLKLSRIQLGL | 676 |
| Db | 549 | E-----DRKT---TVBAIRKMGRTSKVAKHA---ARMQCF | 581 |
| QY | 677 | SKTYAINTLEPHQIRHHKTDLLSPSGTGEVMDGVRMSRSVAKRIRDLVLGSD-VPSAV | 735 |
| Db | 582 | SSTYATVTRPDEVDESDDVVH---NEVIFSDGIGKITPDLALEVAERLQLTDPNPSAY | 638 |
| QY | 736 | QGRFSGAKGMWIDV-DOTGDEDWLETYPSQRKWCDFVDKHQRTLEVRVASELKSAGL | 794 |
| Db | 639 | QIRFAGFKG-VIAVQMGHGDGTRFLRPSMRKFSNHL-----VLEVVS-WTKPQPGFL | 690 |
| QY | 795 | NLQLFVLDEDRADKVMKQCAIGDLINDLQRFSEOKHALNPVEFRQWYVESYSSRAT | 854 |
| Db | 691 | NQIIILLSS-----LNVPSDFWQMQLTSLN---LNNILSDRDVAFLVITSCA | 738 |
| QY | 855 | RVSHGRVPLAGLPSQSEETLNLNMGSPDPKQKYLQDIADWLQKRCDTLKSILNRV | 914 |
| Db | 739 | -----DDGNTAALMLSAGEPEPTEPHLKAMLLAIRSAQLQDLLEKARIFV | 783 |
| QY | 915 | GRSAVYMIADFWGLEE-----NEVHVGFSKFRDEBESFTLLSDCDVLVAR | 962 |
| Db | 784 | PKGRWLMGCLDLVLQEQCFIRATVPSLNSVYFKHGRFSSTDKNTEVILG-TVVIK | 842 |
| QY | 963 | SPAHPSPDQORVAVKPELHSLKDVTPSTGVDPLAKLSGGYDGDMAWVWDPEIV | 1022 |
| Db | 843 | NFLPHGPDVRIIEAVDVPHELHVLVCLVFPQGERPHANEASSGLDGLDYFTWDEKLI | 902 |
| QY | 1023 | DGFVNAEMPLE---PDLRYLKDKTTFKQLMASHGTGSAKEQTYDMIOKSPHFALQP | 1079 |
| Db | 903 | PGKKSNNPMDYSPEA-----KQL-----PRQVSOHLIDFFLKNMISE | 942 |
| QY | 1080 | NFLGMCNTYKRLCYINNVSNNKPAIILSSLVGNLVDSQKQIVFNEASWQLRRELLG | 1139 |
| Db | 943 | NLCRICNAHVHADLSEYGAMDEKICHLAELAATAVDFPKTG-----KLAIMP | 991 |
| QY | 1140 | ALS---LPDPM-----YKSDSWLGR-----GEPTHIIDYLKFSIARP-----A | 1174 |
| Db | 992 | HLKPKVYDFMKEDQSQYSKILGLRYSRTOEASNG-----DVSQEVCTPNDLPD | 1045 |
| QY | 1175 | IDKLEAFHNAKAAKTDGHAHFDWPDJLASYYTTFKEISDKSRSSALLFTLKNRIGEV | 1234 |
| Db | 1046 | IDLEVPGASDFLASAWQCKS---YDAQLSALLSQV-----RVRTAEALVT-----GHI | 1091 |
| QY | 1235 | EKEYGLVKN-KEMEDSKOPVVRVNOVYKWCALTPEAMDKS-GANYDSKVILLELS | 1291 |
| Db | 1092 | T-----FLVKNSSKKQGDIND---RLKTAI-----SALKRKPSTFESIASQCE-- | 1133 |
| QY | 1292 | FLADREMTWALLRASTAPKLYHKSPKFWQWAGRLAYIAQMTSRFEGCAPALMTAF | 1351 |
| Db | 1134 | IGDDEKNLLYEMKASAWQVTVH---PKWVKSRG-----ILGPDGEBIPASLS-- | 1179 |
| QY | 1352 | MYAGLMPDKFKTKQYVARLE 1371 | |
| Db | 1180 | -FAWIPVD-----YLARIK 1192 | |

RESULT 6

Q9ZRX3

ID Q9ZRX3

AC Q9ZRX3

DT 01-MAY-1999

DT 01-MAY-1999

DT 01-JUN-2003

DE RNA-directed

GN RDRP.

OS Petunia hybrida

OC Eukaryota;

OC Spermatophyta;

OC Magnoliophyta;

OC Eudicotyledons;

OC Tracheophyta;

| | |
|----|---|
| OC | lamids; Solanales; Solanaceae; Petunia. |
| OX | NCBI_TaxID=4102; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | TISSUE=leaf; |
| RA | Schiebel W., Pelissier T., Riedel L., Thalmeir S., Schiebel R., |
| RA | Kempe D., Lottspeich F., Saenger H.L., Wassenegeger M., |
| RT | "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from |
| RT | Tomato."; |
| RL | Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases. |
| DR | EMBL; AJ011979; CAA09896.1; - |
| DR | GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA. |
| DR | InterPro; IPR007855; RDRP. |
| DR | Pfam; PF05183; RDRP; 1. |
| KW | RNA-directed RNA polymerase. |
| FT | NON TER 1 |
| FT | NON TER 775 |
| SQ | SEQUENCE 775 AA; 88778 MW; DE43FE50C6A45B93 CRC64; |

Query Match

Best Local Similarity

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

Matches 174; Conservative 119; Mismatches 311; Indels 217; Gaps 32;

Score 322.5; DB 10; Length 775;

Pred No. 3.8e-13;

```
Db 669 YKSEY-----DKKLNLMNDYGIKT-----EAETLGGIMKASITFDRRDAE 711
QY 1284 VIRLLELFLADRENTW-----ALLRASTAFKLYH 1315
Db 712 AIGVAVSL--RKEARTWFKRSDIDDLAKASAWHYVAYH 750

RESULT 7
Q7XM31 PRELIMINARY; PRT; 1136 AA.
AC Q7XM31,
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE OSUNB0048502.13 protein.
GN OSUNB0048502.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.O., Guan J.P., Yu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen L.J., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL606553; CAE04828.1; -
SQ SEQUENCE 1136 AA; 126924 MW; 308B542CCB45AB09 CRC64;

Query Match 4.3%; Score 320.5; DB 10; Length 1136;
Best Local Similarity 22.9%; Pred. No. 9.7e-13;
Matches 160; Conservative 115; Mismatches 248; Indels 175; Gaps 32;

QY 698 LSPSGTGEVNDGVGRMSRAKRDVILGLG--DVPASVQGRFSGAKGMVIDVDVDDGD 755
Db 539 ITDGSKYFSDGKISLRFKRVAAHVHGLDPTNLPNFAQIRGYKG--VIAIDPSS 596
QY 756 EWIETYPQKWECDVDFKHQRTLEVRSAVELSAGNLQLLVLEDR--RDK--VYM 812
Db 597 ID--LSLRPSMKKFESE-----SRMLNITS--WSKSPQCYVNRBIISLLSTLGRIBIFVAM 649
QY 813 ROAIGRLNDLQROFSEQKHALNRVPEFRQWVYESYSSRATRVSHGRVPLAGLPDSQE 872
Db 650 QQ-----DEMRETEM--ITNKEY-----ALSVLKLGGSET 679
QY 873 ETNLFNMGSDFPKKQYLODIANDLQKCDTLKSKLNIRVGRSAYIYMIADFWGLEE 932
Db 680 KTAVKMLQGYEPSPSPYLSMLKAHQENRLTDIRTRCKIHVPKGRVLIGCLDETGVLEY 739
QY 933 NEVHVGF--SSKFRDEESFTLLSP-----CDVLVARSFAHFPSDIQRVAVKPEL 982
Db 740 GQVYITRTKNSEKQDSNOSFYNDGKTATVVGKVAITNCPHGGDRIVLVDLDPDL 799
QY 983 HSLKDVIIIFSTKGDVPLAKLGGDYDGMVAVCWDPDEIVDGFVNAEMPLEDLSRYLKK 1042
Db 800 VGMVDCLVFPQGERPHNPECSGGDLGDLFTTWDKLIPEKVDTPM----- 847
QY 1043 DKTTFFQLWASHGTGSAKEQTYDMIOKSFHALQPNFLGCTNYKRELVCYNNS---V 1099
Db 848 DYTATPRINDH-----VVTLEIIQHFVDYINDSLGAIST--AHLIHADRSPLKA 897
QY 1100 SNKPAILSLVGNLVDQSKQGVFNESAWOALRELLGGALSLED-----PMYKSDS 1152
Db 898 RSEPCQLATLHSAVDPAKTG-----APAEPRTL--RPREYDFMERWEKPMYISNG 949
QY 1153 WLGRGEPTHIIDYKFSIARPAIDKELEAFHNWKAAKOTEDGAHFWDPDL----- 1203
```

RESULT 8

```
Q872R5 PRELIMINARY; PRT; 1351 AA.
AC Q872R5,
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Related to RNA-directed RNA polymerase.
GN B13B3.100.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OC NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX84762; CAD70515.1; -
DR GO; GO:0016329; F:apoptosis regulator activity; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006915; P:apoptosis; IEA.
DR InterPro; IPR000712; Bcl2 BH.
DR InterPro; IPR007855; RDRP.
DR InterPro; IPR003124; WH2.
DR Pfam; PF05183; RDRP; 1.
DR Pfam; PF02205; WH2; 1.
DR PROSITE; PS01259; BH3; 1.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1351 AA; 151975 MW; 0C784E99CDC66AE6 CRC64;
```

Query Match 4.2%; Score 312; DB 3; Length 1351;

Best Local Similarity 22.9%; Pred. No. 4.9e-12; Indels 350; Gaps 54;

Matches 246; Conservative 126; Mismatches 369; Indels 350; Gaps 54;

```
QY 299 FEAPPSOPREK-----RP-----VDATVEAGH-----LIESPSKGRRT 332
Db 92 YEAEERKRRKAEENTSNNAIPARTPTLADADLDCGHSYIDGNLAFITEWSSSLRV 151
QY 333 KSHIDNLOPLSSSQGTSPTTY-----ESFPSSGGEAIPPEPSRNGLARSESASQVQ 388
Db 152 SAKFAKHDLVITSLSAQSQVSIPIRYIHELWMSNGHVAV-----TL 192
QY 389 VHAAPV--VAARLRINWPKFKWLH--EAPLAVAWEVTR--LFMHCKVDLEDESL-- 436
Db 193 AHAPTFLSPPIPLRPPQPKMELEACDSNHAQVSKFLVYHFKV--SDKHLKQNGSYR 250
QY 437 GLKYDPSWSATVDITDKTLTLDLAFRGKPFPEKPNDFVTAMTGNESKGSVVLSA 496
Db 251 GSDPRTAINGLINEQEMFWVTHYTF-----AIOTTSQDAYAVADR-----IRAA 295
QY 497 VLDVNPDS-----PTAPLYLVKLPMPFEOGCRLTRRFGP----- 532
```

Db 296 LLEYEQRDTLPXSLFLNQLVMSYLHPTTVTLKLAKRLADMFETARRSGORQDPSYDA 355
Qy 533 --DRPEIILPPTSPSPVPPVSKPCGAVREVQLWTMG--QH-SLVGQWRAFFAKD 587
Db 356 FXDLFTIDWPSLSS-----ABELAQFEVEGIEHLKTKERMRGGYTLR 402
Qy 588 AG-----YRKLRFQRAEDPKPIIKERV-----HF-----FAETGITERP 624
Db 403 LNEEIPPGTLKIYALVTPTRIHLGHLEAKNRLKFPPEQDHFRLVQFAEED---GQ 459
Qy 625 DVFKTESVVAPEEVEQTEEK-----YSQMLD-----WL----- 654
Db 460 DUFFNSAV--SMDAIYQR--FDVLNGLSVGGRVTRFLGFSHSLRAHSLMLAAPFIYD 515
Qy 655 --LQDNNWTQ-----PHLKFSRIQLGLSKTYAIMTLEPHQIRHHTKDLT---SP 700
Db 516 GKQLASNIIEDLGFNTMSPARAARICQAFSETPYSVSLYDHGI-----DVIRQD 570
Qy 701 SCTGEVWNGVGMERSVAKRI--RDVLGLGDVPSAVQGFSGAKGMWIDVDVDTGDEWI 759
Db 571 KNERVFSGVGLISOGALEVTHREIPESKGYPCNCLQVRWAGAKGMALDARLTGRQICI 630
Qy 760 ETPYSRKWECDVDRHQHTLEVRSVASELKSAGNLQLLPVLEDRARDKVMQRAIGDR 819
Db 631 R--DSNEK---FRSDEEHLICDMASKPIPLMLNRQMIKILE----- 669
Qy 820 LINDLQRFSEOKHALNRVPERQWVYESYSSRATRVSHGRVPPFLAGLPSQSEETLNL- 878
Db 670 -----MRAPA---QWPLELQEKELQ-----LRAITDNVQNVATEFLK 703
Qy 879 -----MNSGFDPKKQKYLQDIAWDLQKPKCDTLKSLNIRGRSA 919
Db 704 LQCVGDSVHLSQFLKDLDRKN--IDYRQDFURGIVEAVVLELRLKHKARIPVPGVT 761
Qy 920 IYMIADFWGLBENEVHVGFSS---KFRDEBSFTLLSDCDVLRVARSAPHFSDIQRYA 976
Db 762 LFGWMDETGLLGEVYVTFVTDGRFKDPPTA-----GPVVVTRSPALHPGDIQIAHN 815
Qy 977 VKFP--EHSKDKVITFSTKGDVPLAKLSGDYGDMAMWCVDPBIV---DGFVNAEM 1030
Db 816 ATPAGHPURELKNCLVFSQNGERDLPQSLSGGLDGTFTNVWQSVAILRTFAADY 875
Qy 1031 P-LEP-DLSRYLK-KDKTTF--KQLMASHGTGSAKEQTYTMDIOKSFHFALQPNFLGMC 1085
Db 876 PRVEPLKLNKVESKDMADFFVEFMKADHLGVIARVHMLAD----- 917
Qy 1086 TNYKELCYINNSVSNKPAIILSSLVGNLVDSKOGIVPNEASWACIARELLGALSPLD 1145
Db 918 ---ER---NBTLDADCLKLAALHSKAVDFSKGI---HVDITELPR-----P 956
Qy 1146 PMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFNMAKAKOTDGAHFWDPDLAS 1205
Db 957 PMYRPD-FLVNGPDIKHD-----KSTIDMEEQYLR-----QDDDDG-----DTP 996
Qy 1206 YTFPKEISDK 1216
Db 997 RYKYK--SDK 1005

RESULT 9

Q95ZG6 PRELIMINARY; PRT; 1964 AA.
ID Q95ZG6
AC Q95ZG6; (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 25, Last annotation update)
DE Putative RNA dependent RNA polymerase (Fragment).
GN RRPB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN (1)
RP SEQUENCE FROM N.A.

RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA Nellen W.,
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
and dsRNase";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314910; CAC41975.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001201; PAP_25A_core.
DR InterPro; IPR007855; RdRP.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF05183; RdRP; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase; Hydrolase.
FT NON_TER 1 1
SQ SEQUENCE 1964 AA; 227413 MW; B93CA9098675E645 CRC64;

Query Match 4.0%; Score 298; DB 5; Length 1964;
Best Local Similarity 20.0%; Pred. No. 8.4e-11;
Matches 173; Conservative 136; Mismatches 291; Indels 266; Gaps 39;

Qy 509 PLVLKVKLPMFTGCGCLTRRFGDPRFFELLISPTSTSPSPVPSVSKQPCAVEEVIQWL 568
Db 772 PSRTIFIQPTL-QRSRAVRKFGSSNFIWKI-----VNESLEPL----- 810
Qy 569 TMGSHSLVGVQWRAFFAKDAGYKPLREFOLRAEDPKPIIKERVHFFAETGITERPDPVK 628
Db 811 -----QGVNKHPLHENRI-----XPL-----EGGILVGKGIYS 840
Qy 629 TRSVVPAEPEVQRTBFKYSQMLDWL--QLDNN--WQ--PHL-----KLFRIQLGLS 677
Db 841 YAG-----NSNSQLREYSWFSVQNGTHTVTKWISGIEHVDNVRKFFRCIGLMFS 890
Qy 678 KTVAILTEPHQIRHHTKDLSPSGTGEVMDGVMSRVAKRIKRDVLGLGDVPSAVQG 737
Db 891 TTPTVTLPQRI-YRIQDI---TRTHVFTGCGEIGPELAKHNLNFRSTCAYQV 946
Qy 738 RFGSAGKMWIDVDDTDEDEWIETPSORKWECDFVDKQRTLEVRSAVELSKSAGLNQ 797
Db 947 RIGENKGMVLVN--NQAPDPGSIYIRPSWVKFNPIVCGDEHRTLEICSV--STTSRCKLNQ 1004
Qy 798 LLPVLED-RARDKVKRQALGDRINDLQRFSEOKHALNRVPEFQWVYESYSSRATRV 856
Db 1005 VISLLSTLGTQDNV-----FFALQDHYLN---QVACIVNDTINASKQAI 1045
Qy 857 SHGRVFPFAGLPDSQBELNFMNSGFDPKKQKYLQDIAWDLQKRCDTLKSXLNIRVGR 916
Db 1046 EF-----FPDITEGEL-----YQDPYIRRLISLVKLKMERIQQKCHIEIKD 1087
Qy 917 SAVIYMIADFWGLBENEVHVGFSSFRDEE--SFTLLSDCDVLRVARSAPHFSDIQRV 974
Db 1088 SRMLLGVCDDPTNSLPPTNVFVQLEEDDDDDGKYEKVEGLVWVWIKPCTHPGVDVYL 1147
Qy 975 RAVFKPELHSLKDVIIIFSTKGDVPLAKLSGDYGDMAMWCVDPBIVDGVNAEMPLEP 1034
Db 1148 KAVDNLELRLHNLVNFSTKGDVNPFEKISGSLDGDVRFCCYDKSLI----- 1195
Qy 1035 DLSRYLKDKTTPKQLMASHGTGSAKEQTYTMDIOKSFHFALQPNFLG---MCTNYKER 1091
Db 1196 -----GNRSKSETAY-----LGDETVSNNDKKA 1218
Qy 1092 LCYINNSVSNKPAIILSSLVGNLVDSKOGIVPNEASWACIARELLGGA-----LSLPDP 1146
Db 1219 -----NVFNPD-FALSSMYSTNVPERQELGMVN---SHLAISDLFGANHFSIISKE 1267
Qy 1147 MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFNMAKAKOTDGAHFWDPDLAS- 1205
Db 1268 CFKE-----IDYPKTGI-HGTIPKCANVW-----LKTGYFHYMQRENSTR 1307

```

QY 1206 -YY-----TFKETS-DKS-----RSSALLFTTLKNRIGE 1233
Db 1308 VYIQKTMKQVYDQIDQVYIGDFLNLSDKSNLVDYELNLSAKILSYQYKLQVHS 1367
QY 1234 VEKEYGR-----LVKNKEMRD-----SKDPYPRVNOVYKWKCAITPEAMDK 1275
Db 1368 LLRHSAESESIMIGFLDQGFISDKVSKDKIGEMKNDY-IKIQOTFEN-----EFLKE 1420
QY 1276 SGANYDSKVIRLLELSFLADRENTW 1301
Db 1421 FGEQHKENCLLHVRVNI---EKKVSAW 1444

RESULT 10
Q9ZR58 PRELIMINARY; PRT; 1114 AA.
AC Q9ZR58;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48).
GN RDRP.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RUIGERS;
RX MEDLINE=93055198; PubMed=9836747;
RA Schiebel W., Pellissier T., Riedel L., Thalmair S., Schiebel R.,
RA Kempe D., Lottspeich F., Sanger H.L., Wassneger M.;
RT "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT tomato."
RL Plant Cell 10:2087-2102(1998).
DR EMBL: Y10403; CAA71421.1; --
DR PIR: T30819; T30819.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR InterPro: IPR007855; RDRP.
DR Pfam: PF05183; RDRP; 1.
KW Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 1114 AA; 126807 MW; B6D30ACD41DB37FF CRC64;

Query Match 4.0%; Score 297; DB 10; Length 1114;
Best Local Similarity 20.4%; Pred. No. 3.9e-11;
Matches 175; Conservative 133; Mismatches 323; Indels 228; Gaps 32;

QY 562 EEVIQWLTMGQSLVGRQWRAFFKADGKPLREPOLRAEDPKPIIKERVHFFAETGIT 621
Db 386 EDIDNLRV---SFVDEWEKLYSTDL-----LPK-----ASTSGG 418
QY 622 FRPDVFK-----TRSVVPAEPEVQRTBFKYSQMLDNLQLDNNTW---OPHLKL--- 568
Db 419 VRTNIYERILSLRKGFIQDKKFE-FLAFSSQL-----RDSVMFASRPGLTANDI 471
QY 669 -----FSRT-----QLGSKTVAIMTLEPHQIRHHTOLL-----SPSGTGEVNMDCV 711
Db 472 RANWGFDSQIKNVAKYAARLGQSGFSRETL---SVLRHEIEVIPDKVHGYSYVPSDGI 528
QY 712 GRMSRSVAKIRIDVLGLGDVPSAQVQGRFGSAKGMWIDVDDTGDEDEWIETPSQRKWECD 771
Db 529 GKISGDFAHRAVASKGLQYTFSAFQIRYGGYKGVGVDPDSSMK---LSLRKSMGKYESD 585
QY 772 FVDKHQRTLEVRSAVSLKSGAGNLQLLPVLEDRARDKVMQRQAIGDRINDLQROFSEQ 831
Db 586 NI-----KLDVLG-WSKYQCYLNRLQITLLS-----TLGVXDEVLEQKQEAVDQ 630
QY 832 KHALNRPEFRQWYYESYSSRATRVSHGRVFFLAGLPDSQEETLNFAMNSGDPKKQKYL 891
Db 631 LDAILHDSLKAQEALELMS-----PGENTNLLKAMLNCYKPKDAEPPL 673

```

```

QY 892 QDIANDLOKRCCTLKSKLNIRVGRSAYITVMTADFWGLVEENEVHVGFSS---SKFRDEE 947
Db 674 SMLQTRFRASKLLDLRTSRIFIPNGRTWMGCLDESRTLEYGVQVQFTGAGHGFSDDL 733
QY 948 ESF-----TLLSDCVLVNARSAPAHFSDIQRVRAVFKPELHSLKDVILFSTKGDVPL 999
Db 734 HPFNNSRSTNSNFIKGNVYVAKNPCIHPGDIRLVKAVNVRALHHMVDVCVFPQKGRPH 793
QY 1000 AKKLSGGDYDGMNAWCVWDPEIVDGFVNAEMPLEPDLRSYLLKKDKTTTFKQLMASHGTGSA 1059
Db 794 PNECSGDLDDDIYFVCWDQDMIPPRQVQPMETPPAPSIQLDHD----- 837
QY 1060 AKSQTYDMTOKSFHFALQNFQMGCTN---YKERLCYINNSVSNKPAIILSLIVGNLV 1115
Db 838 ---VTIEEVEEYFTNYINDSLGIITANARHVVFADR---EPDNAMSDPCKKAELFSIAV 890
QY 1116 DQSKQGIWFNEASWAOLRRRELLGALSLLPDMVKSDSWLGRGEPETHIIDYLFKSIARPAI 1175
Db 891 DFKTGVPAEIPS-QURPK-----EYDFMDKPD-----KTSYISERVIGKLFKRVK 936
QY 1176 DKLEAFHNAMKAAKOTEDGAHFWDPLASYTFFKEISDKSRSSALLFTTLKNRIGEV- 1234
Db 937 DKA-----PQASSIATFRDVARRSYDADMEVDGFEDEIDEAF 974
QY 1235 --EKEYGLVKNKEMRDSKDPYPRVNOVYKWKCAITPEAMDKSGANYD-----SKV 1284
Db 975 DYKTEY-----DNKIGNLMDYGIKTE-----ABILSGGIMKASKTFDRKDAEALISA 1023
QY 1285 IRLELSFLADRENTW-----ALLRASTAFKLYYHKS-----PKFV 1321
Db 1024 VRLR-----KEARAWFKRRNDIDDMLPKASAWYHVTYHTYMGVNOGLKRAHFISFP 1077
QY 1322 WQVAGRQLAVIKAQMTSRP 1340
Db 1078 WCVDYDQLIQIKDKRNRNP 1096

RESULT 11
Q19285 PRELIMINARY; PRT; 1780 AA.
AC Q19285; Q22232;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F10B5.7 protein.
GN F10B5.7.
OS Caenorhabditis elegans
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sims M.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z66500; CAA91312.1; --
DR EMBL; Z48334; CAA91312.1; JOINED.
DR EMBL; Z48334; CAA88315.1; --
DR EMBL; Z66500; CAA88315.1; JOINED.
DR PIR; T20695; T20695.
DR WormPep; F10B5.7; CE02178.
DR InterPro; IPR007855; RDRP.
DR Pfam; PF05183; RDRP; 1.
SQ SEQUENCE 1780 AA; 203145 MW; 16444F29CEFF902C2 CRC64;

Query Match 4.0%; Score 296.5; DB 5; Length 1780;
Best Local Similarity 20.4%; Pred. No. 9e-11;
Matches 196; Conservative 129; Mismatches 348; Indels 289; Gaps 37;

```

```
QY 468 FPEKPPNDVFTVMTGNFES-----KGSVVLSAVLDYNDPNSPT-----APLYL 512
Db 631 FDERQIDV-VTAFITMYQSRKIQVERLLKGESLDQVGLAKPLPKNCVSAKVAVITVTSRI 689
QY 513 VKLKLMEQECRLTRREGDPRFELLIPSTISPSVPPVVKQPGAVEVIEQWLTMQ 572
Db 690 LLMAEVMVMVN-RVRRREGPD----- 709
QY 573 HSLVGRQWRAFFAKDAGYKRLREFQLRAED-PKPIIKERVHFFAETGITFRPDVFK--- 628
Db 710 YAL-----RCVFRDDNLGELARDPSINNIDHMSNIVTEGIIYLTKNGIQVADRVSFLG 764
QY 629 -----TRSVVPAEVPY-----EORTFQKVSQMLDWLLQLDNNTW 662
Db 765 WNSQMRQOGCYLYAPRVNALTGEVTGVEIRVWMDGFRDAISVPKMSRMGQCFTQA- 823
QY 663 QPHLKLFRSIQGLGSKTYAINTLEPHQIRHKTDLSPSGTGEVWN-----DVGGRMSRS 717
Db 824 QP-----TVSSVKNIIHVENIQVRLERHW-IVEPDIEGGVENKYCFSDGGGRISIK 875
QY 718 VAKRIRDLVGLGVDPSAVQGRFGSAGKMWVID--VDTGDEDWIETVPSQKWCDFVK 775
Db 876 LATHISKILQKEVAFQVRFKFGKILVIDPTIDDIINMPKVFIRKSOQKFGEGGEL 935
QY 776 HORTLEVRSAVELSKSAGLNQLLPVLEDRARDKVMGRQ----- 814
Db 936 QDEYLVVUYKAMP-SPVCLNRPFFITL-----DQVSEKQSSASHRRITNRVHYLERELC 989
QY 815 AIGRLINDLQRFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVFFLAGLPDSQDET 874
Db 990 SLSNLNLINEQ-----AABELVNRNTLAIDW--NAASKTA----- 1022
QY 875 LNFLMNSGDPKPKQKQVLODIADLQKCDTLKSKLNI-----RVGRSAYIYMIADFVGL 930
Db 1023 -----GFEISVDPLIKMLFSIYRNIHHSKAKIFLPSLORS--MGVVDFTGLL 1073
QY 931 ENEVHVGFSSKPRDEESFTLLSDCVLVARSFAHPFSDIQRVAVPKPELHSLKDVII 990
Db 1074 QYGVFIQVSPBIR-QTGNRPILTKGVLTKNPCHVPDVRVFDVAVQWPAHALHVDVV 1132
QY 991 PSTKGDVPLAKLSGGDYDGMWVCMVPEIVDGFVNAEMPLEDPLSLYLKDKTTFKQL 1050
Db 1133 FQHGPRPHPDMDAGSDLDGDEYSIIVDQEMLLDY-NEEAVVFP----- 1175
QY 1051 MASHGTGSAAKQTTIYDIQKSFHALQPNFLGMCNTYKRLCYINNS---VSNKPAIL 1107
Db 1176 --SSSAABEDKEPTTDDVVERFLRY-LOQDSIGRMSH-AHLAYADLHGLFHENCHAL 1230
QY 1108 SSLVGNLVDSQKQIVFNEASWAQLRRELLGALSPLDPMYKSDSWLGRGEPTIIDIYK 1167
Db 1231 KCAVA--VDFFPKSGVPAEPLS-----SFEQCEMTPDYMSGGKPMYSTRLN 1275
QY 1168 FSIARPA--IDKELEAFHNAKAAKDTEDGAHFWDPLASYYTFKEISDKSRSSALLFT 1225
Db 1276 GOLHRAKARKVEVLEEF-----ETRSVVF--- 1299
QY 1226 TLKNRIGEVEYGBL-----VKNKENRDSKDPYVRVQVKEKWCALITPEA 1272
Db 1300 -----EREYDKLICPEDVDVFFGNEIKLVQTLRLDEYVDVMQQLLDEY-GIDEA 1349
QY 1273 MDKSGANYDSKVRILLELSFLADREMTWALLRASTAFKLYHK-----SPKFWQVAGR 1327
Db 1350 SVVSG-----HAASIKRLAGMERDDYSFYHTDKVVELRYEKLAVFRAKFFEEFGGE 1401
QY 1328 QL 1329
Db 1402 EI 1403
PRELIMINARY; PRT; 1107 AA.
RESULT 12
Q9LQV2
AC Q9LQV2;
```

```
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE F10B6.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
RT 1.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006917; AAF79241.1; -.
DR GO; GO:0005853; C:eukaryotic translation elongation factor 1. . .; IEA.
DR GO; GO:0003746; F:translation elongation factor activity; IEA.
DR GO; GO:0006414; P:translational elongation; IEA.
DR InterPro; IPR001326; EFL BD.
DR InterPro; IPR007855; RDRF.
DR Pfam; PF05193; RDRP; 1.
SQ SEQUENCE 1107 AA; 126214 MW; F8FF97691FF82187 CRC64;
Query Match 4.0%; Score 295; DB 10; Length 1107;
Best Local Similarity 21.4%; Pred. No. 5.2e-11;
Matches 131; Conservative 105; Mismatches 240; Indels 136; Gaps 21;
QY 638 PVEQRTFVKVQMLDWLLQLDNNTWPHLKFSLRQLGLSKTYAINTLEPHQIRHKTDL 697
Db 470 PIDRIT---AAHIRAMGDFD-----HIRNVAKYAARLQSGFS--SSRETLNVRSEIEV 519
QY 698 LS-----PSGTGVWMDGVGRMSRSVAKRIRDLVGLGDV-PSAVQGRFGSAGKMWVIDV 751
Db 520 IPDVEIISIGTRVFDSDGIGKISAEPAKVAKCGLTETSPSAFOIRYGYKGVAVDPN 579
QY 752 DTGDEDMWETPSQRKWCDFVDKHQTTLVRSVASELSAGLNQLL-----PVLEDR 805
Db 580 SS-----KKLSLRKSKSFESENTKLDVLAWSKYQPCYNNR 615
QY 806 ARDKVNRQAIQDELINDLQRFSEQKHA-LNRPEFRQWVYESYSSRATRVSHGRVFP 864
Db 616 QLITLLSTLGVTDSVFEKKQREVVDRLDAILTHPLB-----AHEALGLM 659
QY 865 AGLPDSQETFLNFMNSGDPKQKQKYLODIADLQKCDTLKSKLNIRVGSAYTYMIA 924
Db 660 A--PGENTNLKALILCGYKPDAPLPFLSMWLQNFRAKSLLELRKTTRIFISGRSMGCL 717
QY 925 DFVGVLNEEVAVGFSKFRDEEESTLLSDCVLVARSFAHPFSDIQRVAVFKPELHS 984
Db 718 DETRTLEYGVVVQVYSDPMRPGRR---FIITGVVVVAKNPCLHPGDVRVLQAVNPALNH 774
```



```
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX NCBI_TaxID=44689;
RN SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
and dsRNase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314909; CAC41974.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0008026; P:ATP dependent helicase activity; IEA.
DR GO; GO:0016787; P:hydrolase activity; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0004872; P:receptor activity; IEA.
DR GO; GO:0006810; P:transporter activity; IEA.
DR GO; GO:0001410; DEAD.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001201; PAP_25A_Core.
DR InterPro; IPR007855; RdRP_boxC.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF05183; RdRP; 1.
DR SMART; SM00487; HELIC_C; 1.
DR SMART; SM00490; HELIC_C; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW ATP-Binding; Helicase; Hydrolase.
FT NON_TER 1883 1883
SQ SEQUENCE 1883 AA; 217454 MW; 1B1D9B9738B15094 CRC64;

Query Match
Best Local Similarity 4.0%; Score 293.5; DB 5; Length 1883;
Matches 173; Conservative 145; Mismatches 305; Indels 221; Gaps 41;

QY 509 PLYVLKPLMPEQCRTRRFGPRDFEILIPSTSPSPVVPVSKQPGAVEVQWL 568
DB 947 PSTFIQPTL-QRSRAVRKFGSSNFIVVKI-----VNESLEPL----- 985
QY 569 TWGQSLVGRQWRAEFADGAYRKPLREFOLGAEDPKIKERVHFFAETGITRPDPVK 628
DB 986 -----QVNNKHLENRI-----KFL-----EGILVGGEIYS 1015
QY 629 TRSVPAEEFVQRTFVKVQMDLL--QLDNNT---WQ---PHL---KLSRIQLGLS 677
DB 1016 YAG-----NNSQLREYSSWFSNQIHTVKIWSGTEHVDNVRKFFRCIGLMFS 1065
QY 678 KTYAINTLEPHQIRHKHTDLLSPSGTGEWMNDGVRMSRSVAKRIDVLGLDVPASVQ 737
DB 1066 TPIPTVLPQNRH-YRIQDI--TRNTH--FTGCGEIGFELAKHLNENFRPSTCAVQV 1121
QY 738 RFGSAKGMWVIDDGTGDEDMETIYPSQKWB-CDFVDKHQRTLEVRVASSELKAGLNL 796
DB 1122 RLGKNGVLVN-NQAPDPSGIYIRPSVMKFNPIDGDEH-RTLEICSV-STTSRCKLNR 1178
QY 797 QLLPVLVD-RARDKVMQRAIGRLINDLQRFSSQKALNRPVFRWVIESVSSRATR 855
DB 1179 QVLSLLSTLGTQDNV-----FFALQDHYLN---QVAQIVNDTNSKQAI 1219
QY 856 VSHGRVPPLAGLPSDQEEETLNFNMNSGDPKPKOKYLODIAMDQKRCDDTLKSLNIRVG 915
DB 1220 VFF-----FPDITEGEL-----YQDPVIRILLISLYLKKMERQCKHLEIK 1261
QY 916 RSAYIMTADFNGVLEENEVHVGFSKFRDEEB---SFTLLSDCDVLVARSPAFPPSDIQ 972
DB 1262 DSRMLLGVCDPTNSLPNTVVFQLESEEDDDDDGRKYEKVETGLVWIKNPCTHPGDVR 1321
QY 973 RYRAVFKPHELISKVLIPTSGDYPLAKKLSGGYDGDMAVWCWDPREIVDGFVNAEMPL 1032
DB 1032
```

```
1322 YLKAVDNIRLRHURNVLVFSTKGDVFNFKFISGSDLDGDRYFFCYDKSLGNRSESE--- 1378
1033 EPDLSRYL-----KDKTTFKQ---LWASHGIGSAKEQTTIDYDMIOKSEHFALQPNF 1081
1379 -----TAYLVVETVSNNDKXANFNDPFPALSSMYSTNAERQE-----LGKLVH---SH 1423
1082 LGMCTNYKERLCYINNSVSNKFAILLSSLVGNLVDQSKOGIVFNEASNAQLRRELGGAL 1141
1424 LAISDQY-----GANHKYSIQISKCEKFEIDYPKGTGIHGTPKEVNIRLQTVG--- 1471
1142 SLDDPM-----YKSDSWLGR-----GEPHTIIDYL-KSIAAPADDKLEAFHNAM 1186
1472 -YPHYMORENSTRVYQSKKIMQYDQDLVYIGDFLNPISLDKSNLVDGVEIYNLSA 1530
1187 KAAKTEDGAHFWDPLASY-----TFPKEISDKSRSSALLFTLLKNRIGEYKEYGRL 1241
1531 KI-----LYSQYKLVHSLRLHYSAESESIM-----IGFLDQGF--- 1565
1242 VKKEMRD-----SKDPYRVNVQVYKWKALTFEAMDKSGANYDSKVILLESFLADRE 1297
1566 ISDKVSKDKGEMKNDY-IKIQTFEN-----EFLKEFGQHKENCLLIHRVNI--EKK 1616
1298 MNTW 1301
1617 VSAW 1620

RESULT 15
Q8HLK9 PRELIMINARY; PRT; 1107 AA.
AC Q8HLK9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE RNA-dependent RNA polymerase 1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yu D., Fan B., MacParlane S.A., Chen Z.;
RT "Analysis of the involvement of an inducible Arabidopsis RNA-dependent
RNA polymerase in antiviral defense.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY148431; AAN64409.1; -.
DR GO; GO:0005853; C:eukaryotic translation elongation factor 1 . . . ; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0003746; F:translation elongation factor activity; IEA.
DR GO; GO:0006414; P:translational elongation; IEA.
DR InterPro; IPR001326; RF1_BD.
DR InterPro; IPR007855; RdRP.
DR Pfam; PF05183; RdRP; 1.
DR RNA-directed RNA polymerase.
SQ SEQUENCE 1107 AA; 126204 MW; C52CFCA1E5880CB7 CRC64;

Query Match
Best Local Similarity 3.9%; Score 287; DB 10; Length 1107;
Matches 130; Conservative 105; Mismatches 241; Indels 136; Gaps 21;

QY 638 PVQRTEFKVQMDMLQLDNNTWQPHLKFRIQLGSKTYAINTLEPHQIRHKHTDL 697
DB 470 PIDRIT---AAHRAWMGDFD-----HIRNVAKYARLQGQSFSS-SRRETLNRSDEIEV 519
QY 698 LS-----PSTGTGEWMNDGVRMSRSVAKRIDVLGLGDV-PSAVQGRFSGAKGMWIDVD 751
DB 520 IPDVEILISLGRVYFSDGIGKISAEFARKVARKCGLTETSPSAFQIRYGYKGVAVDPN 579
QY 752 DTGDEDMETIYPSQKWB-CDFVDKHQRTLEVRVASSELKAGLNLQLL-----PVLEDR 805
DB 580 SS-----KLSLRKSMKSFESENTKLDVLAWSKYQFCYMAR 615
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 06:26:16 ; Search time 5456.8 Seconds
(without alignments)
7672.409 Million cell updates/sec

Title: US-09-913-878A-2

Perfect score: 7397

Sequence: 1 MNPITPRKRNPFVEIINRL.....YEVLGDDDFDGTGTNGDY 1402

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cg2.1/USPTO.spool/US09913878/runat 29032004 082008 8379/app query.fasta_1.2254
-DB=EST -OPMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODS=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878 @CGN 1.1 4565 -runat 29032004 082008 8379 -NCPUS=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|---------------------|
| 1 | 266 | 3.6 | 2199 | 11 | AY103827 | AY103827 Zea mays |
| 2 | 204 | 2.8 | 883 | 14 | CA765940 | CA765940 AF53-Rpf |
| 3 | 203.5 | 2.8 | 972 | 14 | CF885460 | CF885460 trico82xd |
| 4 | 197.5 | 2.7 | 787 | 14 | CB907399 | CB907399 trico82xd |
| 5 | 191.5 | 2.6 | 577 | 10 | BF096404 | BF096404 EST360431 |
| 6 | 188.5 | 2.5 | 3758 | 11 | BC015736 | BC015736 Homo sapi |
| 7 | 186.5 | 2.5 | 737 | 14 | CF883391 | CF883391 trico88xh |
| 8 | 186.5 | 2.5 | 776 | 14 | CB909520 | CB909520 trico88xh |
| 9 | 184.5 | 2.5 | 954 | 29 | CG445102 | CG445102 OGBM57TV |
| 10 | 183.5 | 2.5 | 641 | 14 | CA257764 | CA257764 SCCFL110 |
| 11 | 175.5 | 2.4 | 582 | 12 | B1246074 | B1246074 IP1_68.D1 |
| 12 | 173 | 2.3 | 622 | 13 | BU965311 | BU965311 sat08g07 |
| 13 | 172 | 2.3 | 761 | 10 | BF275078 | BF275078 GA_Eb002 |
| 14 | 171 | 2.3 | 934 | 29 | CG212111 | CG212111 OG0FJ51TV |
| 15 | 170 | 2.3 | 3267 | 11 | BC036553 | BC036553 Homo sapi |
| 16 | 169.5 | 2.3 | 648 | 14 | CA223511 | CA223511 SCJPL1C0 |
| 17 | 169.5 | 2.3 | 741 | 14 | CF441078 | CF441078 EST1677423 |
| 18 | 169 | 2.3 | 3990 | 11 | BC039311 | BC039311 Homo sapi |
| 19 | 168 | 2.3 | 493 | 13 | BU067573 | BU067573 1626_E08 |
| 20 | 168 | 2.3 | 541 | 13 | BQ827639 | BQ827639 gd58f03.Y |
| 21 | 168 | 2.3 | 677 | 14 | CG281556 | CG281556 GR45_104C |
| 22 | 167.5 | 2.3 | 916 | 29 | CG223171 | CG223171 OGWA14TV |
| 23 | 166 | 2.2 | 722 | 14 | CF134248 | CF134248 WHEA369.F |
| 24 | 166 | 2.2 | 812 | 29 | CG140649 | CG140649 FUIRY65TB |
| 25 | 164 | 2.2 | 570 | 10 | BE433646 | BE433646 EST400175 |
| 26 | 163.5 | 2.2 | 585 | 14 | CF244986 | CF244986 3530_1.8 |
| 27 | 163.5 | 2.2 | 4117 | 11 | BC012130 | BC012130 Homo sapi |
| 28 | 163 | 2.2 | 640 | 10 | BR460205 | BR460205 069G02.Ma |
| 29 | 162.5 | 2.2 | 1891 | 29 | CG756328 | CG756328 P051-3-E0 |
| 30 | 161 | 2.2 | 786 | 29 | CG809201 | CG809201 PSAAE20TF |
| 31 | 161 | 2.2 | 3312 | 11 | AK050985 | AK050985 Mus muscu |
| 32 | 161 | 2.2 | 6166 | 29 | AY417941 | AY417941 Mus muscu |
| 33 | 160.5 | 2.2 | 3423 | 11 | AK084676 | AK084676 Mus muscu |
| 34 | 159.5 | 2.2 | 741 | 13 | BQ997497 | BQ997497 QG316J23 |
| 35 | 158 | 2.1 | 5220 | 29 | AY413592 | AY413592 Homo sapi |
| 36 | 157.5 | 2.1 | 621 | 13 | BU997627 | BU997627 HI08T16r |
| 37 | 156 | 2.1 | 1487 | 12 | BQ23047 | BQ23047 602821859 |
| 38 | 154.5 | 2.1 | 629 | 12 | BQ268864 | BQ268864 BJ288864 |
| 39 | 154 | 2.1 | 781 | 29 | CG938103 | CG938103 MERNF75TR |
| 40 | 154 | 2.1 | 1379 | 10 | BF133062 | BF133062 601645474 |
| 41 | 154 | 2.1 | 1704 | 29 | CG756741 | CG756741 P051-4-F0 |
| 42 | 153.5 | 2.1 | 872 | 29 | CG266238 | CG266238 OGXA241TH |
| 43 | 153.5 | 2.1 | 885 | 28 | BH163571 | BH163571 ENTGB82TF |
| 44 | 152.5 | 2.1 | 893 | 29 | CC699849 | CC699849 OGKAJ24TH |
| 45 | 152 | 2.1 | 2072 | 11 | BC048977 | BC048977 Homo sapi |

ALIGNMENTS

RESULT 1
AY103827
LOCUS
DEFINITION Zea mays PC0135426 mRNA sequence.
ACCESSION AY103827
VERSION AY103827.1 GI:21206905
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 2199)

linear HTC 16-OCT-2002

QY 1167 LysPheSerIleAlaArgProAlaIleAspLysGluLeu-----GluAlaPheHisAsn 1184
 Db 1396 ACGAGGAGTGGCAAGCGCTTATGACACCATTTGATTGTTGATGCTATGAAGAT 1445
 QY 1185 AlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAla 1204
 Db 1446 TACATTACTAGGCTATAGATTCAAGGAGAGTACGATTTCAGGCTGGTAACTTATG 1505
 QY 1205 SerTyrTrpThrPhePheLysGlu-----IleSerAspLysSerArgSerSerAla 1221
 Db 1506 GACCACTATGCAATAAAGTGAAGCTGAGATATATAGTGGATGATATCTTAAGATGCA 1565
 QY 1222 LeuLeuPheThr-----ThrLeuLysAsnArgIleGlyGluValGlu 1235
 Db 1566 AAGAATTTCAACCAAGAGTAGTGGTCTGATGCAATTAAGATGCGGTGAGATCTTTGAGG 1625
 QY 1236 LysGlu-----TyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
 Db 1626 AAAGAAGCTAGGTCTGTTCAATGAGATGAGCAGGAGGATGCGCAAGATGCCATG 1685
 QY 1252 AspPro-----TyrProValArgValAsnGlnValTyr----- 1262
 Db 1686 GAGGCCAAGGCTCTGCTGTGTACCATGTTACTTATCATCAGCAGTACTGGGGCAGCTAC 1745
 QY 1263 -----GluLysTrpCysAlaIleThrProGlu 1271
 Db 1746 AATGAAGGATGATGATGCGCGCATCTTATTAGTTCCTCATGTCGATATGACAAAGCTT 1805
 QY 1272 AlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeu 1288
 Db 1806 GTGGCATCAGCAGGAGGAGNACTCTCTCAGCAGATGGATCGAACTTG 1856

RESULT 2
 CA765940
 LOCUS
 DEFINITION
 AF53-Rpf 09_H23_T7_092.ab1 IRRI Drought Stress Panicle Library
 Oryza sativa (indica cultivar-group) cDNA clone C0003263 5', similar
 to unknown, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.

REFERENCE
 AUTHORS
 Bennett, J., Arumugan, K., Lafitte, R., Wen, J., Rudd, S. and
 Bruskiewich, R.M.
 IRRI Drought Stress Panicle cDNA Library
 Unpublished (2002)

TITLE
 JOURNAL
 COMMENT
 On Dec 2, 2002 this sequence version replaced gi:25995195.
 Contact: Richard Bruskiewich
 Biometrics and Bioinformatics Unit
 International Rice Research Institute
 DAP0 7777, Metro Manila, Philippines
 Tel: +63-2-845-0563
 Fax: +63-2-845-0606
 Email: r.bruskiewich@cgiar.org
 International Rice Information System (IRIS);
 http://www.iris.irri.org: D0203262

Assignment of putative function to the sequence by S. Rudd of the
 Munich Information Center for Protein Sequences
 (http://mips.gsf.de)

Plate: 09 row: H column: 23.
 Location/Qualifiers
 1..883
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="IR64"
 /db_xref="taxon:39946"
 /clone="C0003263"

FEATURES
 source

/tissue_type="Panicles"
 /dev_stages="Flowering"
 /clone_lib="IRRI Drought Stress Panicle Library"
 /note="Vector: pBluescript II SK+; Water stress was
 applied by not watering for 4 consecutive days. Panicles
 were collected from control (well watered) and stressed
 plants at 2 days before heading, at heading, 50% flowering
 and 4 days after 50% flowering."

ORIGIN

Alignment Scores:
 Pred. No.: 3,678-05 Length: 883
 Score: 204.00 Matches: 56
 Percent Similarity: 46.26% Conservative: 43
 Best Local Similarity: 26.17% Mismatches: 99
 Query Match: 2.76% Indels: 16
 DB: 14 Gaps: 3

US-09-913-878A-2 (1-1402) x CA765940 (1-883)

QY 824 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 843
 Db 45 TTAATTCGACAGATTATAATATTGCTATCTCTCACTGAATGTCCAGATTCTATCTTTGG 104
 QY 844 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 863
 Db 105 CAAATGCAAGACCATCTTCTTAACCTCAACATATCTATCATCAGACAGATGTTGCT 164
 QY 864 LeuAlaGlyLeuProAspSer-----GlnGluGluThrLeuAsnPheLeuMetAsn 880
 Db 165 TTTGAGTTTAAACAACCTTCATGCTGCTGATGATGAAACACATCGACATGATGCTCACT 224
 QY 881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
 Db 225 GCTGGCTTTGAACCTTAGAACTGACACCACTTGAAGCAATGCTCTTGGCTATAGGCTT 284
 QY 901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
 Db 285 GCACATTCAGAGTCTTTTGGAAAAAGCAAGGATATTGTGCCAAAGGAAAGTGGTTG 344
 QY 921 TyrMetIleAlaAspPheTrpGlyValLeuGluGlu----- 932
 Db 345 ATGGGCTGCTTGAAGAGTCTGGGGTCTTGAGCAAGGCGAGTCTTTATTCGGCAACA 404
 QY 933 -----AsnGluValHisValGlyPheSerSerLysPheArgAspGluGlu 948
 Db 405 GTTCCATCATTTGATAGTTATTTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAA 464
 QY 949 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 968
 Db 465 AACACAGAGGTCTATTTTGGGA---ACTGTGGTAATAGCAAGAAATCCCTGTCTTCATCCA 521
 QY 969 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 988
 Db 522 GGGATGTCGGATCCTCTGAACAGTTGATGTGCCGAACTGATCATCTGTTGATTGT 581
 QY 989 IleIlePheSerThrLysGlyAspValProLeuAlaLysLeuSerGlyGlyAspTyr 1008
 Db 582 TTGGTGTCTCCCAAGAGGTGAGAGGCCACCATGTTTACAGAGGCATCTCGGAGCGATCT 641
 QY 1009 AspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 1022
 Db 642 GATGGGATCTCTTACTTTGTGACATGGGATGAGAACTTATA 683

RESULT 3
 CF885460
 LOCUS
 DEFINITION
 CF885460
 Hypocrea jecorina cDNA clone trico02xd19, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 CF885460
 GI:38140142
 Hypocrea jecorina (anamorph: Trichoderma reesei)

```

ORGANISM      Hypocrea jecorina
REFERENCE      Eukaryota, Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS       Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
              1 (bases 1 to 972)
TITLE         Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
              Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
              Dean,R.A.
JOURNAL        Analysis of the protein processing and secretion pathways in a
COMMENT        Trichoderma reesei EST dataset
              Unpublished (2003)
              Contact: Ralph A. Dean
              Fungal Genomics Laboratory
              North Carolina State University
              Campus Box 7251, Raleigh, NC 27695, USA
              Tel: 919-513-0020
              Fax: 919-513-0024
              Email: ralph.dean@ncsu.edu
              Seq primer: LT-F1 primer.
              Location/Qualifiers
                1..972
                  /organism="Hypocrea jecorina"
                  /mol_type="mRNA"
                  /strain="QM6a"
                  /db_xref="taxon:51453"
                  /clone="trico82xd19"
                  /dev_stage="mycelia"
                  /clone_lib="T.reesei mycelial culture, Version 6 October
                  2003"
                  /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                  culture grown from 24 hrs to 6 days with varying Carbon
                  and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.:      4,64e-05      Length:      972
Score:          203.50      Matches:     51
Percent Similarity: 48.25%      Conservative: 18
Best Local Similarity: 35.66%      Mismatches: 69
Query Match:     2.75%      Indels:      5
DB:              14      Gaps:          1
US-09-913-878A-2 (1-1402) x CF885460 (1-972)
QY 1248 ArgAspSerLysAspProTyrProValArgValAsnGlnValTyrGluLysTyrCysAla 1267
DB 6 CGGGATTGAATAAAGTTCCCAACAGCGCGAGGTCTTACCAAAATGGCTGGAC 65
QY 1268 IleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeu 1287
DB 66 ATACAAACCGCGCGCGAGCTGTGATGAGCGCGAGTTTCCCGCGCTAGTCCATGACGCA 125
QY 1288 LeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTyrAlaLeuArgAlaSer 1307
DB 126 AACAGGCTTCGCGTTTGATAC-----TGGCAGCTGCTGAAGCTTCG 170
QY 1308 ThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTyrGlnMetAlaGlyArg 1327
DB 171 ACCACTTCAAGAAGATATCACACAGCGCGCCCAAGTTTGTCTGGAACATGGCCGGAAG 230
QY 1328 GlnLeuAlaTyrIleLysAlaGlnMetThrSerArgProGlyGluGlyAlaProAlaLeu 1347
DB 231 CAGCTTGCTTACATCAAGCGATGGCGGAACAGCGCTGGCGGACGATCAAGGTGTT 290
QY 1348 MetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLysGlnTyrVal 1367
DB 291 GTGATTCGCGAGATGGGGGGTCTTCGCGCGGCAAGAGCTCATCACCCTCGCTGGCG 350
QY 1368 AlaArgLeuGluLysAspGlySerGluTyrProAspProGluValTyrGluValLeuGly 1387
DB 351 ATACAGAGGAGGCTCGCGGGATTTCGAGAGTGCACACTGGCTTTGAGAGGAGGTGTTGAG 410
QY 1388 AspAspAsp 1390
|||||

```

```

Db 411 TTCGACGAT 419
RESULT 4
CB907399
LOCUS      787 bp      mRNA      linear      EST 02-JUL-2003
DEFINITION      trico82xd19 T.reesei mycelial culture, Version 3 april Hypocrea
                jecorina cDNA clone trico82xd19, mRNA sequence.
ACCESSION      CB907399
VERSION        CB907399.1 GI:30122057
KEYWORDS       EST.
SOURCE         Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM       Hypocrea jecorina
REFERENCE      1 (bases 1 to 787)
AUTHORS       Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
                Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
                Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
                Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
TITLE         Transcriptional regulation of biomass-degrading enzymes in the
                filamentous fungus Trichoderma reesei
JOURNAL        J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE        22803314
PUBMED         12788920
COMMENT        Contact: Pamela K. Foreman
                Genencor Intl. Road, Palo Alto, CA 94304, USA
                925 Page Mill Road, Palo Alto, CA 94304, USA
                Tel: (650) 846-7635
                Fax: (650) 621-7817
                Email: pforeman@genencor.com
                Seq primer: LT-F1 primer.
                Location/Qualifiers
                  1..787
                    /organism="Hypocrea jecorina"
                    /mol_type="mRNA"
                    /strain="QM6a"
                    /db_xref="taxon:51453"
                    /clone="trico82xd19"
                    /dev_stage="mycelia"
                    /clone_lib="T.reesei mycelial culture, Version 3 april"
                    /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                    culture grown from 24 hrs to 6 days with varying Carbon
                    and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.:      9.7e-05      Length:      787
Score:          197.50      Matches:     48
Percent Similarity: 48.51%      Conservative: 17
Best Local Similarity: 35.82%      Mismatches: 64
Query Match:     2.67%      Indels:      5
DB:              14      Gaps:          1
US-09-913-878A-2 (1-1402) x CB907399 (1-787)
QY 1257 ArgValAsnGlnValTyrGluLysTyrCysAlaIleThrProGluAlaMetAspLysSer 1276
DB 110 AAGCGCGAGGCTGTCTACCAAAATGGCTGGACATCAACCGCGCGCGAGCTGTGATG 169
QY 1277 GlyAlaAsnTyrAspSerLysValIleArgLeuGluLeuSerPheLeuAlaAspArg 1296
DB 170 AGCGCGAGTTTCCCGCGCTAGTCCATGAGCAACAGCGCTTCGGCGTTGAGTAAC--- 226
QY 1297 GluMetAsnThrTyrAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLys 1316
DB 227 -----TGGCAGCTGCTGAAGCTTCGACCACTTCAGGAAGATATCACCACAGC 274
QY 1317 SerProLysPheValTyrGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMet 1336
DB 275 GCGCCCAAGTTTCTTGGACATGGCGGAAGAGCTTGCCTACATCAAGCGCGATGGCG 334
QY 1337 ThrSerArgProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeu 1356
|||||

```

Db 335 CGAAGCAGCGCTGGCAGACGTCAAAGCTGCTTGTGATCCCGAGATGTGGGGGCTCCTG 394
 QY 1357 MetProAspLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGlu 1376
 Db 395 CGGCGGACAAAGAGCTCATCACCTCGCTCGCGATACAGAGGAGGCTGCGCGGATCG 454
 QY 1377 TyrProAspProGluValTyrGluValLeuGlyAspAsp 1390
 Db 455 GAGATGCACCTGGCTTGGAGAGGCTGTTGAGTTGACGAT 496
 RESULT 5
 BF096404
 LOCUS
 DEFINITION
 EST360431 tomato nutrient deficient roots Lycopersicon esculentum
 cDNA clone cLEW1106 5' sequence similar to RNA-directed RNA
 polymerase (Arabidopsis thaliana), mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Lycopersicon esculentum (tomato)
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 577)
 van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,P.,
 Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
 Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
 and Tanksley,S.D.
 Generation of ESTs from tomato nutrient-deficient roots
 Unpublished (1999)
 Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers
 1..577
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="TA492"
 /db_xref="taxon:4081"
 /clone="cLEW1106"
 /tissue_type="roots"
 /dev_stage="5-6 weeks old"
 /lab_host="SOLR"
 /clone_lib="tomato nutrient deficient roots"
 /note="Vector: pBluescriptSKCudapt; Site 1: 5' EcoRI;
 site 2: 3' XhoI; Roots were harvested from plants grown
 under the following deficiencies/stresses: 10 mM Al, Zn,
 P, K, Fe, N. mRNA was isolated from individual treatments.
 Proportional aliquots of mRNA of each treatment were mixed
 and used for library construction."
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.000174 Length: 577
 Score: 191.50 Matches: 43
 Percent Similarity: 50.88% Conservative: 44
 Best Local Similarity: 25.15% Mismatches: 71
 Query Match: 2.59% Indels: 13
 DB: 10 Gaps: 2
 US-09-913-878a-2 (1-1402) x BF096404 (1-577)
 QY 874 ThrLeuAsnPheLeuValMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 893
 Db 51 ACAGCGGCTATAATGTGTAGTCAGGGTTAAACCTCAAGTGAACCTATTGAGAGG 110
 QY 894 IleAlaTrpAspGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArg 913
 Db 111 ATGTTGGCTAGCATTAGACTGCTCAGCTTGGCGACCTCAGGAATAAGACAGGATGTTT 170

QY 914 ValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTyrGlyValLeuGluAsn 933
 Db 171 GTTACTTACGGAAGGTGGTGTGATGGCTGTTGGATGAATTAGCTGAACCTTCAGCAAGGC 230
 QY 934 GluValHisValGlyPheSerSer 941
 Db 231 CAATGCTTTATTCAAGTGTCAAGCCCTCTTTGGAGACCTGTTTGTTAAGCATGTCCTCA 290
 QY 942 LysPheArgAspGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 961
 Db 291 GAGTTTCTGAATCAAGAAATAATCTCAAGTAGTAAGAGGCCTT--GTTGTAATTGCA 347
 QY 962 ArgSerProAlaHisPheProSerSerPheGlnArgValArgAlaValPheLysProGlu 981
 Db 348 AAGAACCCCTGCTTTCATCCCGGATGTCAGAGATTCTGGAGGCTGTAGATGTTCTCGT 407
 QY 982 LeuHisSerLeuLysAspValIlePheSerThrLysGlyAspValProLeuAlaLys 1001
 Db 408 TTACACCATCTCTATGATGTGTGCTCTCCACGAAGGGGATAGCCACATCAAT 467
 QY 1002 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1021
 Db 468 GAAGCATCAGGAGTGACCTTGATGTGATCTCTACTTGTGACTTGGATGAAATCTC 527
 QY 1022 ValAspGlyPheValAsnAlaGluMetProLeu 1032
 Db 528 ATTCCACCCAGTAGAAGAAAGCTGGATGCCAATG 560
 RESULT 6
 BC015736
 LOCUS
 DEFINITION
 Homo sapiens transforming, acidic coiled-coil containing protein 2,
 mRNA (cDNA clone IMAGE:4865697), with apparent retained intron.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 3758)
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Schuler,G.D.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Loquellano,N.A., Peters,G.J.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Murray,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Greenchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Schenker,J., Myers,R.M.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butterfield,Y.S., Krzywicki,M.I., Skalska,U., Smalusz,D.E.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22386257
 12477932
 2 (bases 1 to 3758)
 Strausberg,R.
 Direct Submission
 Submitted (09-OCT-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk

Email: cgapbs@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline
 Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Nataesja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 34 Row: 1 Column: 24
 This clone has the following problem: retained intron.

FEATURES

Location/Qualifiers
 1...3758
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4865697"
 /tissue_type="Colon, adenocarcinoma"
 /clone_lib="NIH MGC_15"
 /lab_host="DH10B-R"
 /note="Vector: pORF7"

ORIGIN

Alignment Scores:
 Pred. No.: 0.00516 Length: 3758
 Score: 188.50 Matches: 194
 Percent Similarity: 34.7% Conservative: 141
 Best Local Similarity: 20.1% Misses: 338
 Query Match: 2.55% Indels: 293
 DB: 11 Gaps: 42

US-09-913-878A-2 (1-1402) x BC015736 (1-3758)

| | | | |
|----|-----|---|-----|
| QY | 82 | GUAlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThrLeu | 101 |
| DB | 374 | CAAGCTCCGAGCTCCACCCACCACCCCGGAGTCAT | 415 |
| QY | 102 | ProTyrSerLysGluProArgAlaAlaThrAlaGlnGlnTrpAlaLeuGlnThr | 121 |
| DB | 416 | -----CCAGA-----ACCGAGGTCAGCACACAGCC----- | 442 |
| QY | 122 | ValLeuLeuGluValLeuAsnArgPheMetProProAsnAsnThrProGlyArgThr | 141 |
| DB | 443 | -----ACCCCGGAAGAAC----- | 457 |
| QY | 142 | PheGlyArgThrLeuSerGly-----Pro-SerGlyLeuSerArgProThrSerThrAs | 159 |
| DB | 458 | AGTAGAGACTGGGTTTACCATGATGACCCAGCTGGTCTCAACTCTGACCTCAAGT-- | 515 |
| QY | 159 | nThrLysArgLysAspGluProAlaAsn----- | 168 |
| DB | 516 | -----GATCTGCTGCTCGCTCCCAAGTGTGGGATTACAGGATGTGG | 562 |
| QY | 169 | -----ValThrPheAlaAspProProLysArg-----SerLeuThrArgSerAlaTh | 184 |
| DB | 563 | TTCTGAGACGTCCCTGCTGCTGATGGCCACGGACGACTGGTGGAGGAAGTCCCTT | 622 |
| QY | 184 | rGlyProToileHisGlyAlaAlaAlaLeuProLeuLysPheProAspProValAsnThrG | 204 |
| DB | 623 | CCGTCCCGCTGACACTCTCTCTGCTGCTGCTGATGAAGACAGCCGATA---GCCAG | 679 |
| QY | 204 | ySerLysArgProSerLeuGluSerGluAsnLeuAsn----- | 216 |
| DB | 680 | CAGTGGGACTTACAACTTGGACTTTGACACATCTTGAGCTGTGGATACCTTTTCAGACCTT | 739 |

| | | | |
|----|------|--|------|
| QY | 217 | -----GlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValaAla | 232 |
| DB | 740 | GGAGCTCTGCTGCTCAGACGTAAAGTACAGAGGGCAAGTG-----AACACAGGAG | 793 |
| QY | 232 | aAlaAlaAlaProProValProileAla---SerAlaLeuAspLysValProThrArgAr | 251 |
| DB | 794 | GAAGTCCACGGATTCCGTCCTCATCTAACTTACACTGTCCCGTGTGCTCAGCTGCA | 853 |
| QY | 251 | gHisAlaAsnThrArgAspProThrAlaThrGlyHisArgAlaAspGlnValAspSe | 271 |
| DB | 854 | AGCAGTACCTTTGATGGTCTTCTTCACAGCACTCCGAGGCTGGCCCTTGGCCCT | 913 |
| QY | 271 | rPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSerAlaCysArg---- | 289 |
| DB | 914 | AGATGCATATAGCAGGGTTCAGCAGTCTCTAGTACCTTAAGCGAACTAAAAACC | 973 |
| QY | 290 | -----HisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSe | 304 |
| DB | 974 | GAGCCGCTCTCTTAAAAAAGAAACAGACCACCAAGAAACCCACAGAGACCCCCCAGT | 1033 |
| QY | 304 | rGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuileG | 324 |
| DB | 1034 | GAAGGAGACGCACAGGACCCAGATGAAGAGCCTGTGCCAGTGGGGAGAACTAGC | 1093 |
| QY | 324 | uSerProSerLysGlyArgThrThrLysSerHis-----IleAspAs | 338 |
| DB | 1094 | ATCTGAGACGAAACGGAATCTGCCAAGACGGAAGTCTTACCCAGCCTTATGGAGGA | 1153 |
| QY | 338 | nGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPh | 358 |
| DB | 1154 | GAGCCCTTGAG----- | 1166 |
| QY | 358 | eProSerSerGlyGlyGluAlaIleProGluProSerArgSerAsnGlyLeuAlaAr | 378 |
| DB | 1167 | -CCGCTGGGGGCCCAAGCTGCTGCTCTGGAAGTGTAGAGGGGTGT | 1225 |
| QY | 378 | gSerGluGluSerAlaArgSerGlnValGlnValHisAlaProValValaAlaArgLe | 398 |
| DB | 1226 | CCCCCGGCTTCTGGAGGTGGAGTGCAGAACTACCCCTGTGCGGAGAAACGCT | 1285 |
| QY | 398 | uArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValaTr | 418 |
| DB | 1286 | G-----CCTCTTACCACGGCCCC----- | 1303 |
| QY | 418 | pGluValThrArgLeuPheMetHisCysLysValAspLeuGluAsp----- | 433 |
| DB | 1304 | GGAGGCGAGGAGTAAACCCATCGGATAGCGGGGGCAAGAGGACTCTCCAGCCAAAG | 1363 |
| QY | 434 | -GluSerLeuGlyLeuLysTyrAsp-----ProSerTyrSerThrAlaAr | 448 |
| DB | 1364 | GCACTCGTAAGCTGGAGTTTGACTATTCTGAGGACAAGAGTAGTGGGACAACCCAGCA | 1423 |
| QY | 448 | gAsp----- | 449 |
| DB | 1424 | GGAAAAACCCCTCTACCAAAAAGATAGCAAAAGCCAGTTGCCAAAATGCCCTGAG | 1483 |
| QY | 450 | -ValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPhePr | 469 |
| DB | 1484 | GAGGCCAAAGATGAAAGACACCCCGAAGAACTTGACAACACTCTCTGCTCACCT--CC | 1540 |
| QY | 469 | oGluLysPro-----ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSe | 487 |
| DB | 1541 | CAGATCCCTGCTGAACCCCAATGACATCCCATTTGCTAAAGGTACTTACACCTTTGAT-- | 1598 |
| QY | 487 | rLysGlySerAlaValValLeuSerAlaValLeuAspTyr-----AsnProAspAsnSe | 505 |
| DB | 1599 | -----ATTGACAAGTGGATGACCCCAATTITAA | 1627 |
| QY | 505 | rProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysAr | 525 |
| DB | 1628 | CCCTTTTCTTCCACCTCAAAAATGCAGGAGTCTCCCAAACTCCCAACAA----- | 1679 |

```

Qy 525 gLeuThrArgArgPheGlyProAspArgPheGluLeuLeuPro----- 541
Db 1680 ----TCATACAACTTTGACCCAGACACCTGTGATGAGTCCGTGACCCCTTTAAGACATC 1735
Qy 542 -SerProThrSerThrSerProSerValProProValValSerlySerGlnProGlyAlaVa 561
Db 1736 CTCTAAGACCCCGAGCTCACTCTTAATCCCGAGCCTCTTTGAGATCCAGCAGTGC 1795
Qy 561 lGluGluVal-lleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpA 581
Db 1796 TATGGAACCAATGGAGTGGAGCGGGATGGCTAAACAAGCCCGCCCAAGAAGAAGAC 1855
Qy 581 rGala-----PhePheAlaLysAspAlaGlyTyrArgLysProLeuArg 596
Db 1856 CCCC-CTAAAGACTGACACATTTAGGGTGAAGAAAGTCGCCAAACGGTCTCTCTCTG 1914
Qy 596 luPheGlnLeuArgAlaGluAspPro-----LysProIleI 608
Db 1915 AT-----CCACCTTCCAGAGACCCACCCCGAGCTGCTACACCAGAAACCCACGATGA 1968
Qy 608 leLysGluArgValHisPhePheAlaGluThrGlyIle----- 620
Db 1969 TCTCTCGCGTGTCCACGCCACAGATGAGGAAAGCTGGCGTCAACCAACAGAAAGTGA 2028
Qy 620 ----- 620
Db 2029 CGTGATGACAGTGGACCTAGAGGTGACAAACAGGACTACCCCGACCCCTCGACCTGT 2088
Qy 621 --ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValG 640
Db 2089 CCACCTTT-----GTAAACGAGACCAATTCAGTTTCACTCCACCTGAGGAG---TTGG 2136
Qy 640 luGlnArgThrGluPheLysValSerGlnMet----- 650
Db 2137 ATTACAGAAACTCTTATGAATTTGAATATATGGAGAAATTTGGCTCTCTCTTACCTCAGG 2196
Qy 651 -----LeuAspTrpLeuGlnLeuAspAsnThrTrpGlnProHisL 666
Db 2197 ACGAGATGCCCGAAGACGAGCGCTTGTACCTTATGTTTGACACTTCTCAGGAGAGCC 2256
Qy 666 euLysLeuPheSerArgIleGlnLeuGlyLeuSer----- 677
Db 2257 CTGTCAAGTCATCTCCCTCGCATCTGAGAGTCCCGAGCCGCTGTTCAGGGTCAAGTT 2316
Qy 678 ----LysThrTyrAlaIleMetThr----- 684
Db 2317 TTGAAGACTCAAGCCCTTGTGAACACTGTGCGMAAACACGACATCTCTCCACGAG 2376
Qy 685 --LeuGluProHisGlnIleArgHisLysThrAspLeuLeuSerProSerGlyThrG 704
Db 2377 GACTGGCCCTTAACCAAGAGTCACAC-----TTGACGGTGCAGAGAAATCCT 2424
Qy 704 lyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArgA 724
Db 2425 CCCAAGAGAGTGGAGCCATGGCTTGGGACCCCTTCAAGACGATTAATTAAG 2484
Qy 724 spValLeuGlyLeuGlyAspVal-----ProSerAlaValGlnGlyArgPheGlyS 741
Db 2485 AGGCTGCTCACCCAAACAGACGCTCTCCATCTCCAAAACAGCCCTTGTACTCCCGCATCGGA 2544
Qy 741 erAlaLys-----GlyMetTrpVal-----IleAspValAspAspThrG 754
Db 2545 CCGCTGAGGTGGAGAAACCTGAGCGCTTCTGTTCACGACCGCCGACCTGACCTCTGCC 2604
Qy 754 lyAspGluAspTrpIleGluThrTyrProSerGlnArg-----LysTrpGlyCysA 771
Db 2605 TCCAGATCCCGACAGACAGATCATATAACCAAGGAGAGAGAGGTCTCAGATGGAAGATA 2664
Qy 771 spPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLys 791
Db 2665 AATATGAAGAAACGAGCGCGAAGTGCATGGAATGAGGAAATAGTCCCGCATGAGA 2724
Qy 791 erAlaGlyLeuAsnLeuGlnLeuProValLeuAspArgAlaAspLys----- 809

```

```

Db 2725 AGACC-----ATCGCTCAGATGATAGAGGACGAACACAGAGAGAAGTCTAG 2769
Qy 810 -----ValLysMetArgGlnAlaIleGlyAsp----- 818
Db 2770 TCTCCACACAGCGGTGCAGCAGCTGTTCTGGAGAGGACGAACCCCTGGCCGACCTGA 2829
Qy 819 -----ArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaL 835
Db 2830 ACTCCGTGGAGAGTCTTGGCCGACCTCTTTCAGAGATATAGAGAGATGAAGGAGTCC 2889
Qy 835 eu-----AsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerA 852
Db 2890 TAGAAGGCTTCGCAAGATGAGAGGTCTTGAAGAGATGTGCGCAGGAGTACCTGTCC 2949
Qy 852 rGalaThrArg 855
Db 2950 GGGTGAAGAAG 2960

RESULT 7
LOCUS CF883391
DEFINITION Hypocrea jecorina cDNA clone trico88xh09, mRNA sequence.
ACCESSION CF883391.1 GI:38138073
VERSION 1
KEYWORDS EST
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 737)
AUTHORS Diener S.E., Dankmeyer L., Dunn-Coleman N., Houfek T.D.,
Mitchell T.K., van Solingen P., Teunissen P.J.M., Ward M. and
Dean R.A.
TITLE Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
JOURNAL Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: Lt-F1 primer.
FEATURES
source
1..737
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/db_xref="taxon:51453"
/clone="trico88xh09"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.: 0.00611 Length: 737
Score: 186.50 Matches: 50
Percent Similarity: 46.58% Conservative: 18
Best Local Similarity: 34.25% Mismatches: 71
Query Match: 2.52% Indels: 7
DB: 14 Gaps: 2

US-09-913-878A-2 (1-1402) x CF883391 (1-737)
Qy 1247 MetArgAsp-----SerLysAspProTyrProValArgValAsnGlnValTyrGluLys 1264

```

Db 80 ATGCGGACGGCATTGAAACTCAAAGTTTGCCAAAGCGCGAGCTGTCTACCAAAA 139
 Qy 1265 TrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284
 Db 140 TGCTGGACATCAACCGCGCCGAGCTGCTGATGAGCGGCGAGTTTCCCGCGCTAGTC 199
 Qy 1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304
 Db 200 CATGACCAACAAGGCTTCGGCGTTGAGTAAC-----TGTTTTTGTCTG 244
 Qy 1305 ArgAlaSerThrAlaPheLysLeuTyrHisLysSerProLysPheValTrpGlnMet 1324
 Db 245 AAGCTTCACCACTTCAAGAGTATCACACAGCGGCCCAAGTTTGTCTGGACATG 304
 Qy 1325 AlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgProGlyGluGlyAla 1344
 Db 305 GCGGGAAGCAGCTTCCCTACATCAAGCGGATGGCGGAAACGACGCTGGGAGACGTCA 364
 Qy 1345 ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLys 1364
 Db 365 AAGTGCTGTGATTCGAGAGTGGGGGTCCTCGCGCCGACAGAGCTCATCACC 424
 Qy 1365 GlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspProGluValTyrGlu 1384
 Db 425 TCCTCGCGATACAGAGGAGGCTGCGCGGATTCCGAGATCGACTGGCTTTGGAGGAG 484
 Qy 1385 ValLeuGlyAspAspAsp 1390
 Db 485 GTGTTTGTGAGTTGACGAT 502

RESULT 8
 CB909520 776 bp mRNA linear EST 02-JUL-2003
 LOCUS jecorina cDNA clone trico88xh09, Version 3 april Hypocrea
 DEFINITION

ACCESSION CB909520
 VERSION CB909520.1 GI:30124181
 KEYWORDS EST.
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
 ORGANISM Hypocrea jecorina
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 776)
 Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
 Dunn-Coleman,N.S., Goedegebuur,F., Houfex,T.D., England,G.J.,
 Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
 Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
 Transcriptional regulation of biomass-degrading enzymes in the
 filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)

22803314

12788920

Genencor Intl.

Contact: Pamela K. Foreman

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genencor.com

Seq primer: LT-F1 primer.

Location/Qualifiers

1..776

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clones="trico88xh09"

/dev_stage="mycelia"

/clone_lib="T reesei mycelial culture, Version 3 april"

/note=Vector: pREP3y; Site 1: Not 1/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

ORIGIN

FEATURES

source

1..954

/organism="Zea mays"

/mol_type="genomic DNA"

Alignment Scores:

Pred. No.: 0.000661 Length: 776

Score: 185.50 Matches: 50

Percent Similarity: 46.58% Conservative: 18

Best Local Similarity: 34.25% Mismatches: 71

Query Match: 2.52% Indels: 7

DB: 14 Gaps: 2

US-09-913-878A-2 (1-1402) x CB909520 (1-776)

Qy 1247 MetArgAsp-----SerLysAspProTyrProValArgValAsnGlnValTyrGluLys 1264

Db 80 ATGCGGACGGCATTGAAACTCAAAGTTTGCCAAAGCGCGAGCTGTCTACCAAAA 139

Qy 1265 TrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284

Db 140 TGCTGGACATCAACCGCGCCGAGCTGCTGATGAGCGGCGAGTTTCCCGCGCTAGTC 199

Qy 1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304

Db 200 CATGACCAACAAGGCTTCGGCGTTGAGTAAC-----TGTTTTTGTCTG 244

Qy 1305 ArgAlaSerThrAlaPheLysLeuTyrHisLysSerProLysPheValTrpGlnMet 1324

Db 245 AAGCTTCGACCACTTCAAGAGTATCACACAGCGGCCCAAGTTTGTCTGGACATG 304

Qy 1325 AlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgProGlyGluGlyAla 1344

Db 305 GCGGGAAGCAGCTTCCCTACATCAAGCGGATGGCGGAAACGACGCTGGGAGACGTCA 364

Qy 1345 ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLys 1364

Db 365 AAGTGCTGTGATTCGAGAGTGGGGGTCCTCGCGCCGACAGAGCTCATCACC 424

Qy 1365 GlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspProGluValTyrGlu 1384

Db 425 TCCTCGCGATACAGAGGAGGCTGCGCGGATTCCGAGATCGACTGGCTTTGGAGGAG 484

Qy 1385 ValLeuGlyAspAspAsp 1390

Db 485 GTGTTTGTGAGTTGACGAT 502

RESULT 9

CG445102/c 954 bp DNA linear GSS 17-SEP-2003

LOCUS OGBAH57TV ZM_0.7_1.5 KB Zea mays genomic clone ZMMEMa0850U18,

DEFINITION genomic survey sequence.

ACCESSION CG445102

VERSION CG445102.1 GI:34827337

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 954)

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,

Citek,R.W., Numberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0288

Email: whitelaw@cigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

1..954

/organism="Zea mays"

/mol_type="genomic DNA"

FEATURES

source

1..954

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMB0850J18"
/clone_lib="ZM 0.7 1.5_KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 0.00129 Length: 954
Score: 184.50 Matches: 61
Percent Similarity: 41.51% Conservative: 49
Best Local Similarity: 23.02% Mismatches: 91
Query Match: 2.48% Indels: 65
DB: 29 Gaps: 8

US-09-913-878A-2 (1-1402) x CG445102 (1-954)

QY 871 GlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyr 890
DB 818 CAAGGAACACCGCAGGGCTGATGCTGGCGCTGCTTGGCTCTCGACTGAGCCACAC 759
QY 891 LeuGlnAspIleAla-TTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLe 910
DB 758 CTGAGAGGATGCTACTGGCCATCTGCACCTCGCAGATGCGGGCCTTCTCAACAAACG 699
QY 910 uasIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValle 930
DB 698 TGG-ATCTTTGTGCCAAGGAAGTGCTGATGGCTGCTCGACGATTTGGGATCCT 640
QY 930 uGluGluAsnGluVal-----HisValGly----- 938
DB 639 TGAGCAGGGCGAGTGCTTTCTCGGGTCTCGACGCCATCATCGCGCAGCGCTCGTGAA 580
QY 939 -----PheSerLysPheArgAspGluGluSerPheThrLeuLeuSe 954
DB 579 CGTTCTCCGATATTTCCCTCGAGATTAAGGGCGATCCCATCGATCGACAGGTATCTGG 520
QY 954 rAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVa 974
DB 519 T-----ACGGTTGTGATGGGAGAACCCATGCTTCATCCAGGGGACGTCCTATCT 466
QY 974 lArgAlaValPheLysProGluLeuHisSerLeuLysAspValIlePheSerThrLy 994
DB 465 TGAACCTGTGGATGTCCTCGATCTATATCATCTGCGACTGCTTCGCTTCCCAAGAA 406
QY 994 sGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTr 1014
DB 405 AGTGAGGCGCCACACGCCGCGAGAGCTTCGGGAGCGACCTCGACGCGCGACTACTT 346
QY 1014 pValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu-- 1033
DB 345 CGTGACGTGGGACGAGAGCTTTGTACCTCCCGGAGAGAGAGCTGCACCCCGATGGACTA 286
QY 1034 ----ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAl 1052
DB 285 CTCCTCCCGGTAAACACAGCACTTCCAGCGAC----- 252
QY 1052 aSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSe 1072
DB 251 -----GTACACCAACCAAGTAAGTACACA 229
QY 1072 rPheHisPheAlaLeuGlnProAsnPhe---LeuGlyMetCysThrAsn----- 1087
DB 228 ATACCACTTTGCT-----TCTTCCTTTGTTTATCTGCTGCCATATATACATCAGTTA 175
QY 1088 -----TyrLysGluArgLe 1092
DB 174 ATGGATTTCGGGTGCTCGTGTGTGAACAGGATACCATGATTGTTCTACTTGGAGAGCAT 115
QY 1092 uCysTyrIleAsn 1096
DB 114 GGTATACGACAC 102

RESULT 10
CA257764
LOCUS
DEFINITION
SCCCFL1100F01.9 FL1 Saccharum officinarum cDNA clone SCCFL1100F01
S', mRNA sequence.
CA257764
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
SACCHARUM OFFICINARUM
SACCHARUM OFFICINARUM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE
1 (bases 1 to 641)
Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unicamp.br
Plate: 100 row: F column: 01
Seq primer: T7 Promoter Primer.
Location/Qualifiers
1..641
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCFL1100F01"
/lab_host="DH10B"
/clone_lib="FL1"
/note="Organ: Inflorescence at beginning of development
An unidirectional cDNA library generated from
[Inflorescence at beginning of development (1cm-long)].
cDNA was prepared from polyA+ mRNA using SuperScript
Plasmid System Kit (Invitrogen). The double-strand cDNAs
were fractionated in a sepharose CL-2B 40cm-columns and
fragments sizing between 0.8 and 1.5 Kb were
directionally cloned into the vector. Details of each
source of RNA and library construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"

ORIGIN
Alignment Scores:
Pred. No.: 0.000838 Length: 641
Score: 183.50 Matches: 44
Percent Similarity: 50.30% Conservative: 39
Best Local Similarity: 26.67% Mismatches: 67
Query Match: 2.48% Indels: 15
DB: 14 Gaps: 3

US-09-913-878A-2 (1-1402) x CA257764 (1-641)

QY 871 GlnGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyr 890
DB 25 CAAGGAACACCTCGACACTGATGTTGAGTGTGCTGGCATTTCCACTGGAACACACAC 84
QY 891 LeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeu 910
DB 85 CTGAAGCAAGTGTGTAGTACTTAAGTCTCTCAGCTACTAGGCTTTTGGAGAAACA 144
QY 911 AsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValleu 930
DB 145 AGGATTTTGTGCCAAGGGAGGTGTTGATGGTGTGCTTGTGATGAACCTTGGGATCCTT 204

| | | | |
|--|---|--|-----------------|
| QY | 931 | GlUGluAeNGluValHisValGlyPheSerSer | ----- 941 |
| DB | 205 | GAGCAAGACAGTGTCTTTATCCGGCGTCATCTCCATCACTCAATAATTGCTAGTAAG | 264 |
| QY | 942 | -----LysPhe--ArgAspGluGluSerPheThrLeuLeuSerAspCysAsp | 957 |
| DB | 265 | CATGGACCAAGATTTTCCAAAGCAACAAATAATGCAGAGACCATGTGGGT-----ACT | 318 |
| QY | 958 | ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal | 977 |
| DB | 319 | ATTGTAATGGCAAGAAATCCATGCTTCATCCAGGGAGTCCCGGATTCITGAAGCTGTT | 378 |
| QY | 978 | PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal | 997 |
| DB | 379 | GATGGCTGACCTGACATGCATCTTGTGATTGCTGGTCTTCCCTAAGAAAGGTGATAGG | 438 |
| QY | 998 | ProLeuAlaLysLysLeuSerGlyAspTyrAspGlyAspMetAlaTrpValCysTrp | 1017 |
| DB | 439 | CCACATGCCAATGAAGCATCTGGAGTGATCTGTGAGGGATCTCTACTTTGTGACATGG | 498 |
| QY | 1018 | AspProGluIleVal 1022 | |
| DB | 499 | GATGAACACCTTATA 513 | |
| RESULT 11 | | | |
| LOCUS | B1246074 | | |
| DEFINITION | IP1.68.D10.b1.A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA, mRNA sequence. | 582 bp mRNA linear | EST 17-JUL-2001 |
| ACCESSION | B1246074.1 | GI:14824093 | |
| VERSION | | | |
| KEYWORDS | Sorghum bicolor | | |
| SOURCE | Sorghum bicolor (sorghum) | | |
| ORGANISM | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum. | | |
| REFERENCE | 1 (bases 1 to 582); Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt,L.H. | | |
| AUTHORS | An EST database from sorghum: developing preanthesis pannicles Unpublished (2001) | | |
| TITLE | Contact: Cordonnier-Pratt MM | | |
| JOURNAL | Laboratory for Genomics and Bioinformatics | | |
| COMMENT | The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTwix or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV High quality sequence stop: 560 POLYA=No. | | |
| FEATURES | | | |
| source | Location/Qualifiers | | |
| | 1..582 | | |
| | /organism="Sorghum bicolor" | | |
| | /mol_type="mRNA" | | |
| | /cultivar="BTx623" | | |
| | /db_xref="taxon:4558" | | |
| | /clone_lib="Immature pannicle 1 (IP1)" | | |
| | /note="Organ: Developing preanthesis pannicles; Vector: pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision." | | |
| ORIGIN | | | |
| Alignment Scores: | | | |
| Pred. No.: | 0.00297 | Length: | 582 |
| Score: | 175.50 | Matches: | 42 |
| Percent Similarity: | 49.69% | Conservative: | 38 |
| Best Local Similarity: | 26.09% | Mismatches: | 60 |
| Query Match: | 2.37% | Indels: | 21 |
| DB: | 12 | Gaps: | 3 |
| US-09-913-878A-2 (1-1402) x B1246074 (1-582) | | | |
| QY | 878 | LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp | 897 |
| DB | 17 | ATGTTGAGTGTGGCATTTTCACCTGAACTGACACACCTGAAAGCAATGCTGTAGCT | 76 |
| QY | 898 | LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer | 917 |
| DB | 77 | ATAAGTCTCTACAGCTACTCGGTCTTTTGGAGAGACACAGGATTTTGTGCGAAGGG | 136 |
| QY | 918 | AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisVal | 937 |
| DB | 137 | AGTGTGTTGATGGCTGCCTTGTATGAACCTGGGATCTTGGACGAAGGACAGTCTTTATC | 196 |
| QY | 938 | -----GlyPheSerSer | 941 |
| DB | 197 | CGGGCGTCTATCTCCATCACTCAATAATTCTGTGGTAAGCATGGACCAAGATTTTCTCA | 256 |
| QY | 942 | LysPheArgAspGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla | 961 |
| DB | 257 | GCAACAAAAATGCAGAG-----ACCATTTGGGT-----ACTATCGTAATGGCA | 301 |
| QY | 962 | ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu | 981 |
| DB | 302 | AAGATCCATGCCCTTCATCCAGGGAGTCCCGATCTTGAAGCTGTTGATGTGCTCGAA | 361 |
| QY | 982 | LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys | 1001 |
| DB | 362 | CTGCATCACTGTTGTTGATGCTTGGTCTTCCCCAAGAAAGTGAGAGGCCACATGCCAAT | 421 |
| QY | 1002 | LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle | 1021 |
| DB | 422 | GAAGCATCTGGAGTGTCTTGTGAGGGATCTCTACTTTGTGACATGGATGAAACCTT | 481 |
| QY | 1022 | Val 1022 | |
| DB | 482 | ATA 484 | |
| RESULT 12 | | | |
| LOCUS | BU965311 | 622 bp mRNA linear | EST 21-OCT-2002 |
| DEFINITION | sat08g07.y1.Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1036-13357.5, similar to TR:Q9ZRS8 Q9ZRS8 RNA-DIRECTED RNA POLYMERASE ;, mRNA sequence. | | |
| ACCESSION | BU965311 | GI:24206058 | |
| VERSION | | | |
| KEYWORDS | EST. | | |
| SOURCE | Glycine max (soybean) | | |
| ORGANISM | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine. | | |
| REFERENCE | 1 (bases 1 to 622) | | |
| AUTHORS | Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Pearson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. | | |
| TITLE | Contact: Shoemaker R/Public Soybean EST Project | | |
| JOURNAL | Public Soybean EST Project | | |
| COMMENT | Unpublished (1999) Public Soybean EST Project Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 | | |

Email: estowatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 445.

FEATURES

Location/Qualifiers
1..622
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl036-13357"
/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-cl036"
/notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies pSuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli ElectroMax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

ORIGIN

Alignment Scores:
Pred. No.: 0.0051 Length: 622
Score: 173.00 Matches: 44
Percent Similarity: 50.00% Conservative: 34
Best Local Similarity: 28.21% Mismatches: 70
Query Match: 2.34% Indels: 8
DB: 13 Gaps: 2

US-09-913-878A-2 (1-1402) x BU965311 (1-622)

Qy 875 LeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIle 894
Db 149 CTGAGGAGATGCTCATTTGGTATACAGCTTAATGAGAACCATCTCTTCAATGAG 208
Qy 895 AlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal 914
Db 209 CTTCAAAACATTAGGGCATCAAAACCTTTGGAAATTCGACTTAAATCTAGGATCTTTATT 268
Qy 915 GlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGlu 934
Db 269 CCAAAAGGAGAGCAATGATGGGATGCTAGTGAACCTAGAACCTTAGATATGTCAA 328
Qy 935 ValHisValGlyPheSerSer-----LysPheArgAspGluGluSerPhe 950
Db 329 GTATTGTTCATGTTCTCTAACATPAGCTGCAGAACTCTATCTGATGATTTTTTTTCATAT 388
Qy 951 ThrLeu-----LeuSerAspCysAspValLeuValAlaArgSerProAlaHis 966
Db 389 GATTTCGCAAGAAATTATGCTTAAAGGTAAAGTAGTGTGTAGCAAAAACCCCTGCTTG 448
Qy 967 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 996
Db 449 CACCAGGTGATGTCGCTGTTTACAGCTGTGGATGTCGACATTTGTACCATCGGTG 508
Qy 987 AspValIlellePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 1006
Db 509 GACTGTGTGTTTTCCTTCAAAAAGGACCAAGACCTCATCAATGAGTGTTCGGGAAGT 568
Qy 1007 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 1022

Db

569 GATCTGGATGAGATATCTACTTTGTTGTTGGACCATGAATGATT 616
RESULT 13
BF275078
LOCUS
DEFINITION
GA_EB0023B06f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_EB0023B06f, mRNA sequence.
ACCESSION
BF275078
VERSION
BF275078.1 GI:11206148
KEYWORDS
EST.
SOURCE
Gossypium arboreum
ORGANISM
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
1 (bases 1 to 761)
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
CONTACT: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAAATCAGCTCACTATAGG
High quality sequence stop: 751.
Location/Qualifiers
1..761
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_EB0023B06f"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES

source
1..761
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_EB0023B06f"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Alignment Scores:
Pred. No.: 0.00828 Length: 761
Score: 172.00 Matches: 66
Percent Similarity: 40.14% Conservative: 46
Best Local Similarity: 23.66% Mismatches: 102
Query Match: 2.33% Indels: 65
DB: 10 Gaps: 9
US-09-913-878A-2 (1-1402) x BF275078 (1-761)
Qy 941 SerLysPheArgAspGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
Db 36 TCTAGTTTGTCTGAACCAAGAAAATTTGGAAAGTAATTAAGGGCTG---GTTGTCTA 92
Qy 961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
Db 93 GCTAGATCTCTGCTTCCTCACCTGGAGATATAAGATTTCTAGAAGCAGTTGATGCCCT 152
Qy 981 GluLeuHisSerLeuLysAspValIlellePheSerThrLysGlyAspValProLeuAla 1000
Db 153 GGTTTACACCATTTTGTATGACTGTCTGCTCTCCCTCAAAAGGGTGAAGACCCCATACG 212
Qy 1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
Db 213 AATGAAGCTTCTGGAGTGTCTTGATGGGACCTTTATTGTCACATGGAGAGGCT 272
Qy 1021 lleValAspGlyPheValAsnAlaGluMetProLeu-----GluProAspLeuSerArg 1038

```

Db 273 CTTATCCCTCCCTAGACAGAGAGCTCAACCCATGCGATGACCTGATGAGCTCGA 332
Qy TyrLeuLysLysAspLysThr-----ThrPheLysGlnLeuMetAla 1052
Db 333 GAGTTGAATCCGCCAGTCACCTATAAGGACATAATTGAGTTTTTTTCAAAAACATGGG 392
Qy SerHisGlyThrGlySerAlaAlaLysGluGlnThrThrAspMetIleGlnLysSer 1072
Db 393 AATAGCACCTGGGAAGTATCTGCAATGACATGTGGTTCACTTCACTGACCTAGTAA---- 449
Qy PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeu 1092
Db 449 ----- 449
Qy CysTyrIleAsnAsnSerValSerAsnLysProAlaIleLeuSerSerLeuValGly 1112
Db 450 -----CAITGGTCTTCAGATGAGAAATGATATCATCTCCAGAGTTAGCGGT 497
Qy AsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeu 1132
Db 498 ATAGCTGTGATTTTCTTAAACCGGAAGATTGTGCGATGCT-----GTCAAATTA 551
Qy ArgArgGluLeuGlyGlyAlaLeuSerLeuProAspProMet----- 1147
Db 552 AAACCAAACTT-----TACCAGATTTTATGGTTAAGGAAGAGGTTC 593
Qy TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTy 1165
Db 594 CAGTCATTTACAAAGTCAAAACAAATTCCTGGAAGA----- 627
Qy rLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAl 1185
Db 628 -----CTATATCGTTATATCAGATGCTTATGACCAGATGCTCTGAAATCTTCTGAGCT 663
Qy aMetLysAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu 1203
Db 684 CAATTTTGGTCTAGTGACATC-----ATTATGATGCGGATCTT 723

RESULT 14
LOCUS CG212111 934 bp DNA linear GSS 22-AUG-2003
DEFINITION G0FJ51TV ZM_0.7_1.5_KB Zea mays genomic clone ZM85Ma070506,
genomic survey sequence.
ACCESSION CG212111
VERSION CG212111.1 GI:34111941
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 934)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: G0FJ51TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5943
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..934
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM85Ma070506"

```

```

/clone lib="ZM_0.7_1.5_KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 0.0135 Length: 934
Score: 171.00 Matches: 67
Percent Similarity: 37.84% Conservative: 45
Best Local Similarity: 22.64% Mismatches: 84
Query Match: 2.31% Indels: 100
DB: 29 Gaps: 12

US-09-913-878A-2 (1-1402) x CG212111 (1-934)
Qy 925 AspPheTrpGlyValLeuGluAsnGluVal-----HisVal 937
Db 27 GAGCAGTTTGGATCTTGGACGAGCGCATGCTTTCTCCGGTCTCGAGCCATCATGTC 86
Qy 938 Gly-----PheSerSerLysPheArgAspGluGlu 948
Db 87 GGCAGCGCTGCTGAACCGTTCTCCGATATTTCCTCGAGTATTAAGGCGATGCCGAT 146
Qy 949 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 968
Db 147 GCACAGTGATTAAGTGT-----ACGGTGTGATGCGAAGAACCCATGCTTATCCA 200
Qy 969 SerAspIleGlnArgValAlaValPheLysProGluLeuHisSerLeuLysAspVal 988
Db 201 GGGCAGCTCGTATCTTGAAGCTGTGGATGCTCTGATCTGATCACCTTGTCCAGCTGC 260
Qy 989 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 1008
Db 261 TTGCTCTTCCCCAAGAAAGTGAGAGCCACACGCGCAGAGCTTCCGGGAGCGACCTC 320
Qy 1009 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAlaAspGlyPheValAsnAla 1028
Db 321 GACGGCAGCTCTACTTCGTGACGTGGACGAGAGCGCTTGTACCTCCCGGAGGAGAGC 380
Qy 1029 GluMetProLeuGlu-----ProAspLeuSerArgTyrLeuLysLysAspLysThrThr 1046
Db 381 TGCACCCCGATGAGCTACTCCCGGTAAACAAAGCAACTTCCACGGCAGC----- 431
Qy 1047 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr 1066
Db 432 -----GTACAC 437
Qy 1067 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe-----LeuGlyMetCys 1085
Db 438 CAACAAGTAAGTACACATACCACTTGTCT-----TCTCTCTTGTGTTTCTTCTGCTGC 491
Qy 1086 ThrAsn----- 1087
Db 492 CATATAATACATCAAGTTAATGATTTCTCGCTGCTCGTGTGTTAAACAGGATACCATTCAT 551
Qy 1088 ---TyrLysGluArgLeuCysTyrIleAsn----- 1096
Db 552 TTCTACTTGGAGAGCATGTTATACCAACCTTGGTCGAATATGCAAGCCCATCTGCTGT 611
Qy 1097 -----AsnSerValSerAsnLysProAlaIleIleLeuSerSerLeu 1110
Db 612 CATGCTGATCGGAGCGACGACGAGCGATGATCCCAAGTGCATCGAGTTGGCTGGCTGC 671
Qy 1111 ValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAla 1130
Db 672 GCATCTATCCCGTGGACTCCCGCCACAGACCGGGAAATTTGTAAGA----- 716
Qy 1131 GlnLeuArgGluLeuLeuGlyAlaLeuSer-----LeuProAspProMet 1147
Db 717 -----ATGCCACGAGCCCTTTCCCAAAAGAGTACCCGACTTCATG 758
Qy 1148 -----TyrLysSerAspSerTrpLeuGlyArg 1156

```


Db 759 GGAAGGAGGAGCCCATCTCTTACAGATCCGAAAGAGATCTTGGAGG 806

RESULT 15
BC036553
LOCUS
DEFINITION Homo sapiens cDNA clone IMAGE526565, containing frame-shift errors.
ACCESSION BC036553
VERSION BC036553.1 GI:23306459
KEYWORDS HTO.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3267)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, K.H., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalek, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 3267)

Strausberg, R.

Direct Submission

Submitted (16-AUG-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

Contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Place: 73 Row: m Column: 13

This clone has the following problem: frame shifted.

Location/Qualifiers

1..3267

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:526565"

/tissue_type="Testis"

/clone_lib="NIH MGC_97"

/lab_host="DH10B"

/note="Vector: pBluescript"

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 0.109 Length: 3267
Score: 170.00 Matches: 226
Percent Similarity: 33.98% Conservative: 143
Best Local Similarity: 20.81% Mismatches: 413
Query Match: 2.30% Indels: 306
DB: 11 Gaps: 54

US-09-913-878A-2 (1-1402) x BC036553 (1-3267)

QY 66 TTPATGSLPASPSPSerLeuAenGlnAlaGluAAsnPhePheIleGluAlaLysAla 85
DB 94 TGGAGACAGAGGATCGCGGGAGGGCGGAGACCCCA-----GTCCCGCCACAGGAG 147
QY 86 Ala-----SerSerAsnTrpValProLysAlaHisAlaAspProAspThrLeu 101
DB 148 GCGAAGCGGGGTCTCGGGTAGTTCACCGCCCGCGCAGTCCCGCCCGC----- 198
QY 102 ProTrpSerLysGluProProArgAlaAlaThrAaGlyGlnGlnTrpAlaLeuGlnThr 121
DB 199 -----CAGTCCCGCCCGCGCGCGCGCGCGCACCGCT----- 231
QY 122 ValLeuLeuGluValLeuAAsnArgPheMetProProProAsnAsnThrPro----- 138
DB 232 -----GAAGTCATTGCGCGAGCGCCCGCCCTCGCGCGCCCGCGCTCTGG 282
QY 139 -----GlyArgThrPheGlyArgThrLeuSerGlyProSer----- 150
DB 283 CTCTGCGCGGGAGGGGGGAGAGAACCCGAGCTGCAACTCGCGCTGGAGCTTTGAG 342
QY 151 ---GlyLeuSerArgProThrSerThrAsnThrLysArg---LysAspGluProAlaAsn 168
DB 343 CTGGGCTCAGCCCGGGATCTAGCACCGCTACCGCGCGAGTCCCGCGCGCGG 402
QY 169 ValThr-----Phe 171
DB 403 GCCTCAGCCCGCGGCTTGGCCATGGCAGCGCGCGAGTGGCGGAATCTGGCCCTG 462
QY 172 AlaAspProPro-----LysArgSerLeuThrArgSerAlaThrGlyProPro----- 187
DB 463 CGTCCCGCGCGCTAGCTAGCAGCAGCGTTACAGGCATCCCGCTTCTCTGCGCGT 522
QY 188 IleHisGlyAlaAlaLeuProLeuLysPheProAspProValAsnThrGlySerLys----- 206
DB 523 GTGCGCGAGCGCCACGCTCCGGCAAGTTCCGCGCGCGAA-GAAGCAGGAGGAGCGAGC 581
QY 207 -----ArgProSerLeuGluSerGluAAsnLeuAAsnGlnCysThrLysArgAla 222
DB 582 TGAGGAGTGGGGCTCTCGCGCCACCGCGCGCTCTCGCGCGGATGT----- 629
QY 223 LysGlyLysLeuSer-----AspAsnValAlaAlaAlaAlaProValProIle 240
DB 630 ---GGGGCGCTCTCTGGCTCCGGAGCGCGCGCGGGATAGCCCGCGGAGCGCGCA 686
QY 241 AlaSerAlaLeuAspLysVal----- 247
DB 687 GCTTCCCTCGCGCCGAGATTTATCATCATGCTGCTACGATGCTGCTAATGATCATTTGGC 746
QY 248 ---ProThrArgArgHisAlaAsnThrArgAspProThrAlaThrGlyHisArgAla 266
DB 747 AGCCACACTGTTCATCT-AACCATCGAATACCATATCAGGAGACACAGTATAGCT 805
QY 267 AspGlnValAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSer 286
DB 806 TCT-----GACAGTGGAGACACTGGCACTTGGAACTCTCTGTCTGATAGCGTG----- 853
QY 287 AlaCysArgHisAAsnGlnSerThrGlnSerSerPheGluAlaProProSerGlnPro 306
DB 854 -----GAAGATCATTCACCTTCAGTGGCGACATTCATCTTTTAAAGCTAGTCAATCA 904
QY 307 ArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuLeuGluSerPro 326

Db 905 TTGATTACTCTTCCT-----ACTGCCATGTGATGCCGTCTAAT 943
Qy SerLysGlyArgThrThiLys-----Ser 334
Db TCCAGCGCTTCAATTTCCAACTTAGGAATCATTGACACACAGATGGCTCAAAATGGAGT 1003
Qy 335 HisIleAspAsnGlnProLeuSerSerSerGlnGlyGluThrSerPheSerThrTyr 354
Db 1004 ACAAGCCTCATCAACATTGGAAACACCGTAGGGGGGACGAGCTCTCCCTAGAC 1063
Qy 355 TyrGluSerPhePro-----SerSerGlyGlyGluGluValAlaIleProGlu 369
Db 1064 ATGAGGACTCTCGGCCACTTCGAAATGGTCATCTTTATCCAACTCACTGCCCGCAT 1123
Qy 370 ProSerArgSerAsnGlnLeuAlaArgSerGluGluSerAlaArgSerGlnValGlnVal 389
Db 1124 AACTGTGGCCAGGTGGCTGTATGTCAGAGAAGAGTCAAGAAATGGTGTGGAGNAGATA 1183
Qy 390 ---HisAlaProValValAlaAlaArgLeuArgAsnIleTTPProLysPheProLysTrp 408
Db 1184 GCGAAGGCTAAGGCTTTAACTACATCACAACTGAGACAAATTTGG-----CCTAGCTGT 1234
Qy 409 LeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPheMetHisCysLys 428
Db 1235 TTACATGAT-----AGTATGGAGATGCTTAG-----1261
Qy 429 ValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrpSerThrAlaArg 448
Db 1262 -----TTAGAGGACAAGAAATAAATAAAACGG-----TCATCAACTCTG 1303
Qy 449 AspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPhe 468
Db 1304 GACTGCAAG-----TATAAA-----1318
Qy 469 ProGluLysProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLys 488
Db 1319 -----TTTGAGCTGTAGCAAGGAGGACTTTAGAGCCTCT 1354
Qy 489 GlySerAlaValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAla 508
Db 1355 TCCTCTACTCTTAGGAGACAGACTGTA-----GACATGACATATAGTGCC 1399
Qy 509 ProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArg 528
Db 1400 -----TTACTGAAACAGACCCATTTATGACAAAG-----TCAGAG 1435
Qy 529 ArgPheGlyProAspArgPheGluIleLeuIleProSerProThrSerThrSerPro 548
Db 1436 GCCTTTGAACTCCAAATAT-----TTAATGCTTGGTCAACAGCAGTAGGTGGAGTT 1489
Qy 549 SerValProValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeu 568
Db 1490 CCCATTGAGCTTCGTTAAGACTCAG-----ATGTGGCTT 1525
Qy 569 Thr-----MetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPheAlaLys 586
Db 1526 ACAGAGCAGCTCGGACAAATCTTTGGAAGGTAGAAATACAGAGGATCTTACAGTTTA 1585
Qy 587 AspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysPro 606
Db 1586 GCTCCTTGGCAACAGCAGCAATTTGAAGACTTTTCGACAGAAGTAGAACACCAATGCGAG 1645
Qy 607 IleIleLysGluArgValHisPhePheAlaGluThrGlyIle-----ThrPheArg 623
Db 1646 GTTTTG-----ACTGGATCATCTCGTCAAAATTTATCA 1678
Qy 624 Pro-----AspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
Db 1679 CCTGGCTATCAGGATTTTCAGTAAGTGGGAATCAATGTTGAAATATAAAGGAGGACTCTTA 1738
Qy 641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeu-----657
Db 1739 AGCGAAGAAATTTGTAATCGATCGCGAGAGCAACAAATTAACCCACTGTCATGAGAGG 1798

Qy 658 -----AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGly 675
Db 1799 ATAAGGATATCAATACGGCTCAACATGCCATGTTAGGACATTATGTA-----AAT 1852
Qy 676 LeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisLysThr 695
Db 1853 TGTGAGGATTTCTTAT---GTGGTAGTTTGCAGCCA-----CAATATGAGAACT 1900
Qy 696 AspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSer 715
Db 1901 TCACTCCAGACACATTTTCAGAGGAATCAGITTTCCATCCCAACAGGGAAATTTGAG 1960
Qy 716 ArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaVal 735
Db 1961 CAAAAGCTTCATCTACTACTGAGAAAGAGTTTACAGCTTAATGAGTTTCTCAACAAAGA 2020
Qy 736 GlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAspThrGlyAsp 755
Db 2021 CTAAAGCCTATTAGTGAAGAGAAGAAACTA-----2053
Qy 756 GluAspTrpIleGluThr-----TyrProSerGlnArgLysTrpGluCysAspPhe 772
Db 2054 GAGGAAAACTTAAACTAGAGATCGATACATCAGTAGTCTGAAAGAAATCCAGAAAG 2113
Qy 773 ValAspLysHis-----GlnArgThrLeuGluValArgSer-----784
Db 2114 GAATCTGACAAACAAAGAAAGAGAGAGATTGAGACCTTGGAAAAAGTATCTGGCT 2173
Qy 785 -----ValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuProVal 801
Db 2174 GATCTTCCAACTCTAGATGATGTACAGAGTCAGAGTCACAGCTGCAGATTTCTGGAAGA 2233
Qy 802 LeuGlu-----AspArgAlaArgAspLysValLysMetArgGln 814
Db 2234 AAAAAATTAAGAAATTACAAGAGCTTTGATAGATACAGAAAAAAACTTGAAGAGAT--- 2290
Qy 815 AlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnHisAla 834
Db 2291 -----CAAAAACAGCTGTCACAGATAAAGAGACACA 2320
Qy 835 LeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThr 854
Db 2321 GTTAAATGCCAGAAAAAGAAAGAGAGGATTAGTAACCTACCGTTTCAGAGTTTGCACA 2380
Qy 855 Arg-----ValSerHisGlyArg-ValPropheLe 864
Db 2381 AAAAGTAGAAGATGCCTTGAAGATGGAATCCGCTTTCCCATGTTAGATGCAAAACAGCT 2440
Qy 864 uAlaGlyLeuProAspSer-----GlnGluGluThrLeuAsnPheLeuMetAsnSe 881
Db 2441 TCAGAAATGAAATGATATCTCAGACAAACAAATGAGACTGCTAGTAGAATAATAGACAG 2500
Qy 881 rGlyPheAspProLysLysGlnLysTyrLeu-----GlnAspIleAlaTrpAspLeuGl 899
Db 2501 CCAACAGATGAGATTACAGAAATGATTTAGAATTCAGTCTATGCAAGAAAGCTTTC 2560
Qy 899 nLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTy 919
Db 2561 TAAAGAGAACTGACCACTCAAAAG-----CTGAAAAAGAAAGAAATGTACAGAG 2626
Qy 919 rIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPh 939
Db 2586 -----ATGATGGAAGAG-----2585
Qy 939 eSerSerLysPhe-----ArgAspGluGluGluSerPheThrLeuLeu-- 953
Db 2627 ATTAACAAAAGCATTCGTTGAAAAACCAAGACACAGATGAGACATCTCTCTATTGGA 2686
Qy 954 -----SerAspCysAspValLeuValAlaArgSerProAlaHi 966
Db 2687 TCAGGGCCAGGAACCTGACCAATCTAGCCACGACAGCTTCTTTCCAAACGGCCACTATT 2746

```
QY 966 sPheProSerAspIleGlnArgValArgAlaValPheLysPro----- 980
Db 2747 T-----GATTGACTGTGATTGATCAGCTGTTCAAGGAAATGCTCTGTTGTTTGT 2797
QY 991 -GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAl 1000
Db 2798 TGACTTGAAGCATTGTGTAGTATTCTTAATCAGCGTGCTCAGGGCAAGGAGCCTAATCT 2857
QY 1000 aLysLysLeuSerGly 1005
Db 2858 TTCATTATTCTGGGA 2873
```

Search completed: March 31, 2004, 12:59:51
Job time : 5497.3 secs